GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: October 19, 2002, 06:35:21 ; Search time 59 Seconds
(without alignments)
862.234 Million cell updates/sec

Title: US-09-807-459-2
Perfect score: 2359
Sequence: 1 MAPSDSVGDVTKTLLAASES.....DPSKALIRKVSTEADNLLEK 458

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2359	100.0	458	22	AAB60669	Babesia caballi me
2	996.5	42.2	513	14	AAR39900	21B4/rhoptry prote
3	827.5	35.1	480	16	AAR77249	Babesia merozoite
4	826.5	35.0	564	14	AAR30613	Babesia bovis immu
5	826.5	35.0	565	17	AAR97981	Babesia merozoite
6	770.5	32.7	456	14	AAR39902	B. canis 21B4/rhop
7	759.5	32.2	496	14	AAR39901	21B4/rhoptry gene
8	161	6.8	91	13	AAR25187	21B4 gene clone pr
9	138	5.8	56	13	AAR25188	21B4 gene clone pr
10	121	5.1	974	21	AAY93246	An Escherichia col
11	120	5.1	1038	22	AAG67416	Amino acid sequenc
12	119.5	5.1	1663	15	AAR46608	Plasmodium falcipa
13	118.5	5.0	1588	15	AAR46605	Malarial PfEMP3 ep
14	118.5	5.0	2441	21	AAB18161	Plasmodium falcipa
15	113.5	4.8	1480	22	ABB59227	Drosophila melanog
16	111	4.7	474	20	AAY02371	Polypeptide identi
17	111	4.7	885	18	AAW22230	K. lactis origin o
18	111	4.7	885	18	AAW14136	Kluyveromyces lact
19	110	4.7	566	11	AAR08259	Haemagglutinin. I
20	109.5	4.6	2211	22	ABG06855	Novel human diagno
21	109	4.6	524	20	AAY20011	B. burgdorferi ant
22	109	4.6	553	20	AAY20010	B. burgdorferi ant
23	109	4.6	1054	22	ABG14594	Novel human diagno
24	109	4.6	1104	22	ABG12107	Novel human diagno
25	108	4.6	2470	22	ABB61247	Drosophila melanog
26	107.5	4.6	446	21	AAG53797	Arabidopsis thalia
27	107.5	4.6	522	21	AAG53796	Arabidopsis thalia
28	107.5	4.6	610	21	AAG53795	Arabidopsis thalia
29	106.5	4.5	714	22	AAG79240	Amino acid sequenc
30	106	4.5	1508	22	AAB50676	C. elegans alpha-2
31	106	4.5	1519	22	AAB50677	C. elegans alpha-2
32	105.5	4.5	541	22	ABG08970	Novel human diagno
33	105.5	4.5	572	18	AAW01670	Influenza A/Texas/
34	105.5	4.5	572	20	AAW75442	Influenza virus A/
35	105.5	4.5	572	22	AAE04952	Influenza virus A/
36	105	4.5	465	22	AAU33839	Staphylococcus aur
37	105	4.5	697	20	AAY20015	B. burgdorferi ant
38	105	4.5	716	22	AAU36810	Staphylococcus aur
39	105	4.5	719	20	AAY20014	B. burgdorferi ant
40	105	4.5	719	22	AAG79241	Amino acid sequenc
41	104.5	4.4	1030	18	AAW19602	Mycoplasma pneumon
42	104	4.4	1871	20	AAY01078	Sugarcane bacillif
43	104	4.4	1871	21	AAB15671	Sugarcane bacillif
44	104	4.4	1871	21	AAY57165	Amino acid seqeunc
45	104	4.4	2437	22	AAU34338	Staphylococcus aur

ALIGNMENTS

RESULT 1
AAB60669
ID AAB60669 standard; Protein; 458 AA.
XX
AC AAB60669;
XX
DT 22-MAY-2001 (first entry)
XX
DE Babesia caballi merozoite 48 kD rhoptry protein.
XX
KW Merozoite protein; 48 kD rhoptry protein; antigen; antibody;
KW recombinant production; diagnosis; equine babesiosis;
KW parasitic infection; veterinary.
XX
OS Babesia caballi.
XX
PN WO200112813-A1.
XX
PD 22-FEB-2001.
XX
PF 13-AUG-1999; 99WO-JP04386.
XX
PR 13-AUG-1999; 99WO-JP04386.
XX
PA (KAGA) CHEMO-SERO-THERAPEUTIC RES INST.
PA (MIKA/) MIKAMI T.
XX
PI Mikami T, Ikadai H, Igarashi I, Suzuki N, Nagasawa H, Fujisaki K;
XX
DR WPI; 2001-202867/20.
DR N-PSDB; AAF59961.
XX
PT Gene encoding merozoite protein of Babesia caballi for diagnosis of
PT equine babesiosis caused by this organism -

```
XX
PS   Claim 2; Page 22-24; 27pp; Japanese.
XX
CC   The invention relates to a 48 kD merozoite rhoptry protein from Babesia
CC   caballi (AAB60669) and cDNA encoding it (AAF59961). The invention also
CC   relates to phage vectors containing a nucleic acid encoding the
CC   merozoite protein, a method for the recombinant production of the
CC   protein, an antibody against the protein, and a method for the diagnosis
CC   of equine babesiosis from horse blood samples by using the antibody to
CC   detect Babesia caballi merozoites, or by using the 48 kD protein as an
CC   antigen to detect anti-Babesia caballi antibodies. The 48 kD merozoite
CC   protein, or an antibody specific for the protein may be used for the
CC   diagnosis of equine babesiosis caused by Babesia caballi. The present
CC   sequence represents the Babesia caballi merozoite 48 kD rhoptry protein.
XX
SQ   Sequence   458 AA;

  Query Match            100.0%;  Score 2359;  DB 22;  Length 458;
  Best Local Similarity  100.0%;  Pred. No. 5e-183;
  Matches  458;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MAPSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCS 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MAPSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCS 60

Qy      61 ASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEK 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEK 120

Qy     121 RFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 RFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEF 180

Qy     181 GASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIP 240

Qy     241 ALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREP 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREP 300

Qy     301 SKALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTV 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 SKALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTV 360

Qy     361 DFINNEIRDPSKALIRKVSTGAEDLFENKIGQGTVDFINNEIRDPSKALIRKVYTEADDL 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 DFINNEIRDPSKALIRKVSTGAEDLFENKIGQGTVDFINNEIRDPSKALIRKVYTEADDL 420

Qy     421 FENKIGQGTVDFINKEIRDPSKALIRKVSTEADNLLEK 458
           |||||||||||||||||||||||||||||||||||||||
Db     421 FENKIGQGTVDFINKEIRDPSKALIRKVSTEADNLLEK 458


RESULT 2
AAR39900
ID   AAR39900 standard; Protein; 513 AA.
XX
AC   AAR39900;
XX
DT   13-JAN-1994  (first entry)
XX
DE   21B4/rhoptry protein 1-4 representative sequence.
XX
KW   Polymerase chain reaction; PCR; amplify; primer; detection;
KW   babesiosis; parasite; Babesia bovis; 21B4/rhoptry; antigen; gene;
KW   repeat region; immune response; vaccine.
XX
OS   Synthetic.
XX
PN   WO9314204-A.
XX
PD   22-JUL-1993.
XX
PF   15-JAN-1993;   93WO-AU00012.
XX
PR   15-JAN-1992;   92AU-0000399.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Dalrymple BP,  Peters JM;
XX
DR   WPI; 1993-243219/30.
DR   N-PSDB; AAQ47074.
XX
PT   Detecting closely linked gene copies which encode protective
PT   antigen against babesiosis - by screening babesial genomic DNA
PT   library with oligo-nucleotide probe based partial sequencing of
PT   protective antigen and identifying positive clones
XX
PS   Claim 22; Fig 5; 55pp; English.
XX
CC   This sequence is a protein which is representative of the Babesia bovis
CC   21B4/rhoptry antigen gene region.  The DNA ecoding this sequence was
CC   isolated by PCR using the primers given in AAQ47068-72.   Primer 21B4.1
CC   corresponds to part of the repeated region of 21B4/rhoptry antigen.  In
CC   hybridisation assays this primer recognised two tandemly repeated
CC   regions suggesting that B. bovis contains two copies of the 21B4/
CC   rhoptry antigen gene.  The two proteins encoded by the two antigen
CC   genes are identical.  Primers 21B4.2 and 21B4.3 flank the 21B4-309
CC   coding region of the antigen gene.  Primer 21B4.4 primes synthesis
CC   just 3' to the end of the open reading frame.  The entire open reading
CC   frame was shown to encode five antigen genes.  The 3' non-repetitive
CC   sequences of open reading frames 1-4 are identical.  Gene 5 shows
CC   sequence divergence throughout most of the open reading frame.
CC   Babesia antigen genes can be used in the production of a combined
CC   vaccine which will stimulate a greater immune response and afford
CC   broader immunity than a single antigen vaccine.
XX
SQ   Sequence   513 AA;

  Query Match            42.2%;  Score 996.5;  DB 14;  Length 513;
  Best Local Similarity  45.2%;  Pred. No. 1.9e-72;
  Matches  201;  Conservative   77;  Mismatches  150;  Indels   17;  Gaps    6;

Qy       1 MAPSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCS 60
             :||:: ||:|  ||  | | : |      : :|| | :  : :|| : :|:: |: |||
Db      31 LAPAEVVGNVAATLSTADEIISSHDYDKELINDMRDQMREMGENFVDTVCTKAPEDSNCR 90

Qy      61 ASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEK 120
             |: |  ||    || : :: ||:: :||||:||:||||:||| ||| | :||| :  |
Db      91 QMVALYADRCEMYGCLQIDNVNYPVDEEYQPLSLPNPYQLDAAFTLFKNSASNPAKNGLK 150

Qy     121 RFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEF 180
             ||||| || |  :|  :  || | |||   |||:|   :: ||||| |||||  |:
Db     151 GQWMRFRNGKKHGDYHRFLIGLLAKGVTRKDGATDLEFLVNKLLYMATTYYKTYLIVNNS 210

Qy     181 GASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIP 240
             | |||:::| | :||:|||:||| |:|||:|  :| |||: |: |:   ||||| ||:|
Db     211 DAKFFNRIAFATKIFGFGIKKALKDIVRSNVPEYMG-EHSIERISHLAHGYKDYMLTQVP 269

Qy     241 ALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREP 300
             | |||:|:| ||:: ||:::|||| ||||:|: |:|:| ||::  |||: |:|  : |
Db     270 TLSKFAERYSDMVMKVLLSSLAGYVKAPWYKRWFNKVKDFFVNKIGKPTKEHFHK--KHP 327

Qy     301 SKA-LKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKE-KVSNDAKDLFENKIGQG 358
             |   :|:   ||| |||||  | |||  :|  |:|   | |:    || ||||||
Db     328 RTAEFFDKMHERTKDFFENKIGAPTKDFFENKIGAPTKDFFENKIGAPTKDFFENKIGAP 387

Qy     359 TVDFINNEIRDPSKALIRKVSTGAEDLFENKIGQGTVDFINNEIRDPSKALI-RKVYTEA 417
           | ||  |:|  |:|           | ||||||  | ||  |:: : :|     |:
Db     388 TKDFFENKIGAPTK----------DFFENKIGAPTKDFFENKLPERTKDFFENKLPEHT 436

Qy     418 DDLFENKIGQGTVDFINKEIRDPSK 442
           | |||| : | ||   :|  | |
Db     437 KDFFENKLPEHTKDFFENKIGAPIK 461
```

RESULT 3
AAR77249
ID AAR77249 standard; Protein; 480 AA.
XX
AC AAR77249;
XX
DT 14-NOV-1995 (first entry)
XX
DE Babesia merozoite p58.
XX
KW Merozoite; surface protein; antigen; p58; babesiosis; vaccine.
XX
OS Babesia bigemina.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= Sig_peptide
XX
PN US5422428-A.
XX
PD 06-JUN-1995.
XX
PF 27-MAR-1987; 87US-0031328.
XX
PR 06-DEC-1991; 91US-0803636.
PR 27-MAR-1987; 87US-0031328.
PR 01-MAR-1991; 91US-0663255.
XX
PA (UNIW) UNIV WASHINGTON STATE.
XX
PI Davis WC, McElwain TF, McGuire TC, Perryman LE;
XX
DR WPI; 1995-214706/28.
DR N-PSDB; AAQ90252.
XX
PT Babesia merozoite 45 kD surface protein from B. bigemina - used in
PT vaccines for the prophylaxis of bovine babesiosis.
XX
PS Disclosure; Column 31-34; 30pp; English.
XX
CC Antigenic surface proteins (45, 55 and 58 kDa) were isolated from
CC the intraerythrocytic merozoite stage of B. bigemina JG-29. The 58
CC kDa surface protein (AAR77249) was characterized, and encoding
CC cDNA (AAQ90252) was isolated from a lambda GEM11 library.
XX
SQ Sequence 480 AA;

Query Match 35.1%; Score 827.5; DB 16; Length 480;
Best Local Similarity 41.8%; Pred. No. 8.7e-59;
Matches 181; Conservative 75; Mismatches 140; Indels 37; Gaps 12;

```
Qy       4 SDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCSASV 63
           :: ||||:|||| |:| |::   |  :| ||   || | :     :| :|   | |  ||
Db      33 AEVVGDVSKTLLEANEVVNAEMEATQVNKDMQSQLSNVKETIVGEVCEKVAGNSTCGESV 92

Qy      64 SAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEKRFW 123
            ||::|| : ||||| |:      ||:||:||:||||:|||:|||:|||:|||  :  ||||
Db      93 IAYVNRCDEGDCLTLDSM------KYKPLSLPNPYQLDAAFMLFRESDSNPAKNEVKRFW 146

Qy     124 MRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEFGAS 183
           || |   :|  :|  | :||:||| || :: |:|||||:| || |||||||  ||   |
Db     147 MRSR--SSHGDYHHFVVSLLKKNVVRDPESNDVENFASQYFYMTTLYYKTYLTVDFTAAK 204

Qy     184 FFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALP 243
           |||||:||| |||:||::|||:::|||||:|:|| |  : :: | | | :|| ||:||:
Db     205 FFNKLAFTTRLFGFGIQKALKRLVRSNLPVDLGT-HPEATIREIASGYGEYMMTQVPAMT 263

Qy     244 KFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREPSK- 302
            ||:||| |  : || ||: ||  | ||:|| | | |:||  |    |   | : :| |
Db     264 SFAERFSKMATKTLLVTVSDYVHLPAYKRWYRKFKEFIVN-FFTDPAKLIMKHVSQPVKT 322

Qy     303 ALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKI------G 356
           |  : |  : :    | :|| |    |  :|| |: :||    :: :   |:
Db     323 AYTKLVPEEHRQAIRNVVGQSTKHIAN-GVRDLSRMIKE----PSQQIIREKLPHYLSKA 377

Qy     357 QGTVDFINNEIRD--PSKALIRKVSTGAEDLFENKIGQGTVDFINNEIRDPSKALIRKVY 414
           :| |: :  :::   | |   :|    :|   |  :  |  |    |   | :  :  |
Db     378 KGAVEHVVKKVKSVVPIK---QKGDQPSEAAVEETVPSG--DSAETEFEVPEEQYVDAVT 432

Qy     415 T--------EADD 419
           |        :|||
Db     433 TQEVNSEKVDADD 445
```

RESULT 4
AAR30613
ID AAR30613 standard; Protein; 564 AA.
XX
AC AAR30613;
XX
DT 06-MAY-1993 (first entry)
XX
DE Babesia bovis immunoreactive 60kD merozoite surface epitope.
XX
KW babesiosis; cows; cattle; bos taurus; babesia bovis; babesia bigemina;
KW merozoite; schizont; ss.
XX
OS Babesia bovis.
XX
PN US5171685-A.
XX
PD 15-DEC-1992.
XX
PF 04-APR-1990; 90US-0504461.
XX
PR 04-APR-1990; 90US-0504461.
XX
PA (UYFL) UNIV FLORIDA.
PA (USDA) US SEC OF AGRIC.
XX
PI Davis WC, Goff WL, Hines SA, Jasmer DP, Mcelwain TF;
PI Mcguire TC, Palmergh, Perryman LE, Reduker DW;
XX
DR WPI; 1993-008582/01.
DR N-PSDB; AAQ33064.
XX
PT DNA encoding Babesia bovis protein - is used as probes and for
PT prodn. of polypeptide(s) for use in vaccines and for prodn. of
PT antibodies
XX
PS Example 19; Fig 3; 20pp; English.
XX
CC This sequence is a 60kD immunoreactive epitope located on the
CC surface of babesia bovis merozoites. This sequence was decoded from
CC the DNA isolated as in AAQ33064. It may be used to raise neutralising
CC antibodies, and as such may be used in the formulation of subunit
CC vaccines for bovine babesiosis. Monoclonal antibodies raised
CC against the protein may be used to identify merozoite surface
CC antigens and may be used in the treatment and/or diagnosis of bovine
CC babesiosis.
XX
SQ Sequence 564 AA;

Query Match 35.0%; Score 826.5; DB 14; Length 564;
Best Local Similarity 35.5%; Pred. No. 1.3e-58;
Matches 178; Conservative 79; Mismatches 195; Indels 49; Gaps 6;

```
Qy       1 MAPSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCS 60
           :||:: |||:| ||  |   :    : : |  ||   ||   :     :||  |: |||
Db      30 LAPAEVVGDLTSTLETADTLMTLRDHMHNITKDMKHVLSNGREQIVNDVCSNAPEDSNCR 89

Qy      61 ASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEK 120
            |: |  ||    | |: ::|||| :||||:||:|||||||| ||||| :||| :: |
Db      90 EVVNNYADRCEMYGCFTIDNVKYPLYQEYQPLSLPNPYQLDAAFRLFKESASNPAKNSVK 149
```

```
Qy    121 RFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEF 180
Db    150 REWLRFRNGANHGDYHYFVTGLLNNNVVHEEGTTDVEYLVNKVLYMATMNYKTYLTVNSM 209

Qy    181 GASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIP 240
Db    210 NAKFFNRFSFTTKIFSRRIRQTLSDIIRWNVPEDF-EERSIERITQLTSSYEDYMLTQIP 268

Qy    241 ALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKF-------- 292
Db    269 TLSKFARRYADMVKKVLLGSLTSYVEAPWYKRWIKKFRDFFSKNVTQPTKKFIEDTNEVT 328

Qy    293 -------------------------FNKEIREPSKALKEKVSTDTKDLFENKIGQGTVD 326
Db    329 KNYLKANVAEPTKKFMQDTHEKTKGYLKENVAEPTKTFFKEAPQVTKHFFDENIGQPTKE 388

Qy    327 FFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLF 386
Db    389 FFREAPQATKHFLDENIGQPTKEFF-REAPQATKHFLGENIAQPTKEFFKDVPQVTKKVI 447

Qy    387 ENKIGQGTVDF-----------INNEIRDPSKALIRKVYTEADDLFENKIGQGTVDFINK 435
Db    448 TENIAQPTKEFRREVPHATMKVLNENIAQPAKEIIHEFGTGAKN-FISAAHEGTKQFLNE 506

Qy    436 EIRDPSKALIR-KVSTEADNL 455
Db    507 TVGQPTKEFLNGALETTKDAL 527
```

RESULT 5
AAR97981
ID AAR97981 standard; Protein; 565 AA.
XX
AC AAR97981;
XX
DT 15-OCT-1996 (first entry)
XX
DE Babesia merozoite surface protein Bv60.
XX
KW Babesiosis; merozoite protein; vaccine; monoclonal antibody.
XX
OS Babesia bovis.
XX
PN US5518916-A.
XX
PD 21-MAY-1996.
XX
PF 04-APR-1989; 89US-0333155.
XX
PR 04-APR-1990; 90US-0504461.
PR 04-APR-1989; 89US-0333155.
PR 14-DEC-1992; 92US-0989616.
PR 21-NOV-1994; 94US-0342480.
XX
PA (USDA) US SEC OF AGRIC.
XX
PI Goff WL, Jasmer DP, McElwain TF, McGuire TC, Reduker DW;
PI Stiller D;
XX
DR WPI; 1996-259067/26.
DR N-PSDB; AAT18995.
XX
PT New fragment of Babesia bovis genomic DNA - useful as a probe for
PT detecting Babesia infection
XX
PS Example 18; Fig 4; 19pp; English.
XX
CC A 60 kDa immunoreactive protein (AAR97981), Bv60, is located on the
CC surface of Babesia bovis merozoites. It is the product of a DNA
CC clone (AAT18995) obtd. from a B. bovis blood-stage library by
CC immunoscreening with monospecific anti-Bv460 antisera. Recombinant
CC Bv60, Bo44 and Bv42 (see also AAR97979 and AAR97980) can be used in the
CC formulation of vaccines utilised for the prophylaxis of bovine
CC babesiosis, and to raise monoclonal antibodies useful for detecting
CC Babesia antigens.
XX
SQ Sequence 565 AA;

Query Match 35.0%; Score 826.5; DB 17; Length 565;
Best Local Similarity 35.5%; Pred. No. 1.3e-58;
Matches 178; Conservative 79; Mismatches 195; Indels 49; Gaps 6;

```
Qy      1 MAPSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCS 60
Db     31 LAPAEVVGDLTSTLETADTLMTLRDHMHNITKDMKHVLSNGREQIVNDVCSNAPEDSNCR 90

Qy     61 ASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEK 120
Db     91 EVVNNYADRCEMYGCFTIDNVKYPLYQEYQPLSLPNPYQLDAAFRLFKESASNPAKNSVK 150

Qy    121 RFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEF 180
Db    151 REWLRFRNGANHGDYHYFVTGLLNNNVVHEEGTTDVEYLVNKVLYMATMNYKTYLTVNSM 210

Qy    181 GASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIP 240
Db    211 NAKFFNRFSFTTKIFSRRIRQTLSDIIRWNVPEDF-EERSIERITQLTSSYEDYMLTQIP 269

Qy    241 ALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKF-------- 292
Db    270 TLSKFARRYADMVKKVLLGSLTSYVEAPWYKRWIKKFRDFFSKNVTQPTKKFIEDTNEVT 329

Qy    293 -------------------------FNKEIREPSKALKEKVSTDTKDLFENKIGQGTVD 326
Db    330 KNYLKANVAEPTKKFMQDTHEKTKGYLKENVAEPTKTFFKEAPQVTKHFFDENIGQPTKE 389

Qy    327 FFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLF 386
Db    390 FFREAPQATKHFLDENIGQPTKEFF-REAPQATKHFLGENIAQPTKEFFKDVPQVTKKVI 448

Qy    387 ENKIGQGTVDF-----------INNEIRDPSKALIRKVYTEADDLFENKIGQGTVDFINK 435
Db    449 TENIAQPTKEFRREVPHATMKVLNENIAQPAKEIIHEFGTGAKN-FISAAHEGTKQFLNE 507

Qy    436 EIRDPSKALIR-KVSTEADNL 455
Db    508 TVGQPTKEFLNGALETTKDAL 528
```

RESULT 6
AAR39902
ID AAR39902 standard; Protein; 456 AA.
XX
AC AAR39902;
XX
DT 13-JAN-1994 (first entry)
XX
DE B. canis 21B4/rhoptry antigen 2.
XX
KW Polymerase chain reaction; PCR; amplify; primer; detection;
KW babesiosis; parasite; Babesia bovis; 21B4/rhoptry; antigen; gene;
KW repeat region; immune response; vaccine.
XX
OS Babesia canis.
XX
PN WO9314204-A.
XX
PD 22-JUL-1993.
XX
PF 15-JAN-1993; 93WO-AU00012.
XX
PR 15-JAN-1992; 92AU-0000399.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX

PI Dalrymple BP, Peters JM;
XX
DR WPI; 1993-243219/30.
DR N-PSDB; AAQ47076.
XX
PT Detecting closely linked gene copies which encode protective
PT antigen against babesiosis - by screening babesial genomic DNA
PT library with oligo-nucleotide probe based partial sequencing of
PT protective antigen and identifying positive clones
XX
PS Claim 24; Fig 7; 55pp; English.
XX
CC This sequence is encoded by the Babesia canis 21B4/rhoptry antigen
CC gene 2. The DNA encoding this sequence was determined from restriction
CC fragments from the clone B. canis lambda GEM-11 #9. B. canis was found
CC to contain two genes which are related to the B. bovis 21B4 gene. Gene
CC 1 and gene 2 are very similar but gene 2 appears to contain a large
CC number of repeats. Babesia antigen genes can be used in the production
CC of a combined vaccine which will stimulate a greater immune response
CC and afford broader immunity than a single antigen vaccine. See also
CC AAR39899-901.
XX
SQ Sequence 456 AA;

Query Match 32.7%; Score 770.5; DB 14; Length 456;
Best Local Similarity 35.0%; Pred. No. 3.4e-54;
Matches 161; Conservative 95; Mismatches 161; Indels 43; Gaps 9;

```
Qy       1 MAPSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCS 60
           :: ||       |||    |  :|   | :|: |::: :   :   | :|  : : : |
Db      31 LSKSDGAEKTLSTLLNVDASTRAALEGYRMNAAMANFSNGRREEEEEAVCGNIAEETECQ 90

Qy      61 ASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEK 120
            ||: |:  | : || :::: ||| | :|||||||:||||||| :|: |::||  :  :
Db      91 KSVAEYVESCVRYDCFSIENQKYPQEKEYQPLTLPNPYQLEAAFYVFRNSESNPIKNPTE 150

Qy     121 RFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEF 180
           |||||| |  :  :|: : |:| ||::      ::| |  :| |||||:|||||| :|
Db     151 AFWMRFRHGGRYGAYHNFLVNILYKNLSDSMVDDNLEGFVRKYAYMATMYYKTYTALDVV 210

Qy     181 GASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIP 240
           |   ||::|:  |||  |: ||  |||||:| | | :::| ||:|:   |::||  |:|
Db     211 NARIINKIAFSRHLFGRQIRNALTNIIRSNIPEDFG-KYNVDRLRHVMGGYEEYMMKQVP 269

Qy     241 ALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREP 300
           :|| |||::: |||: |:  |  |   ||:||  :::|| ||::  |||:||  :| ||
Db     270 SLPNFAKKYAGMVVKSLIKNVGAYQKQPWFKKLNNQIRNFFVNKIHEPTKEFFVNKIHEP 329

Qy     301 SKALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTV 360
                  ||: | ||| : | :||  :| :|:           |: | ||: : |
Db     330 ----------TKEFFVNKIHEPTKEFFVNKIHEPT----------KEFFVNKLHEPTK 367

Qy     361 DFINNEIRDPSKALIRKVSTGAEDLFENKIGQGTVDFINNEIRDPSKALIRKVYTEADDL 420
           :|  |:: :|:|     :  ||      | |:         :|  || :: :   |
Db     368 EFFVNKLHEPTKEFFSNMVPGAFQKISEKAGR--------HLRS-SKTVVPE--DEPSSS 416

Qy     421 FENKI---GQGTV-DFINKEIRDP-----SKALIRKVSTE 451
           ||:    || |:  |  : |:  |     |:  : :|  |
Db     417 LENEAVEDGQLTMGDVTDFEMATPTYEQGSQESLNEVGNE 456
```

RESULT 7
AAR39901
ID AAR39901 standard; Protein; 496 AA.
XX
AC AAR39901;
XX
DT 13-JAN-1994 (first entry)
XX
DE 21B4/rhoptry gene 5 antigen.
XX
KW Polymerase chain reaction; PCR; amplify; primer; detection;
KW babesiosis; parasite; Babesia bovis; 21B4/rhoptry; antigen; gene;
KW repeat region; immune response; vaccine.
XX
OS Synthetic.
XX
PN WO9314204-A.
XX
PD 22-JUL-1993.
XX
PF 15-JAN-1993; 93WO-AU00012.
XX
PR 15-JAN-1992; 92AU-0000399.
XX
PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
PI Dalrymple BP, Peters JM;
XX
DR WPI; 1993-243219/30.
DR N-PSDB; AAQ47075.
XX
PT Detecting closely linked gene copies which encode protective
PT antigen against babesiosis - by screening babesial genomic DNA
PT library with oligo-nucleotide probe based partial sequencing of
PT protective antigen and identifying positive clones
XX
PS Claim 26; Fig 6; 55pp; English.
XX
CC This sequence represents the Babesia bovis 21B4/rhoptry antigen
CC encoded by gene 5. The DNA encoding this sequence was isolated
CC by PCR using the primers given in AAQ47068-72. Primer 21B4.1
CC corresponds to part of the repeated region of 21B4/rhoptry antigen.
CC In hybridisation assays this primer recognised two tandemly repeated
CC regions suggesting that B. bovis contains two copies of the 21B4/
CC rhoptry antigen gene. The two proteins encoded by the two antigen
CC genes are identical. Primers 21B4.2 and 21B4.3 flank the 21B4-309
CC coding region of the antigen gene. Primer 21B4.4 primes synthesis
CC just 3' to the end of the open reading frame. The entire open
CC reading frame was shown to encode five antigen genes. The 3'
CC non-repetitive sequences of open reading frames 1-4 are identical.
CC Gene 5 shows sequence divergence throughout most of the open reading
CC frame. Babesia antigen genes can be used in the production of a
CC combined vaccine which will stimulate a greater immune response and
CC afford broader immunity than a single antigen vaccine.
XX
SQ Sequence 496 AA;

Query Match 32.2%; Score 759.5; DB 14; Length 496;
Best Local Similarity 34.3%; Pred. No. 3e-53;
Matches 161; Conservative 93; Mismatches 173; Indels 43; Gaps 8;

```
Qy       1 MAPSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCS 60
           :||:: |||:| ||  | | :::    : || ||   |      | ::|| :| : | |
Db      31 LAPAEVVGDLTHTLKIADEIINAENIEHEINRDMHLRLVEEGSKFIDQICQEVAEDSKCR 90

Qy      61 ASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEK 120
             | :|: || : :|| :  : |||  :|||| ||:||||:||| |||  ::||| :  |
Db      91 EQVESYVKRCEENNCLQIDEVAYPLNQEYQPLLLPEPYQLDAAFTLFKNCESNPAKNGLK 150

Qy     121 RFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEF 180
             |||:: || |  :|  : :|| |:: |    ||:|   :: ||||| ||||| | :|
Db     151 GPWMRYKEGKEHGDYHHFIISLLGKSLVRKDGVTDLEFLVNKLLYMATTYYKTYLIVKKF 210

Qy     181 GASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIP 240
           || |||  |||  :|| ||||||| |:|||:| |:| |||: |: |::  |||| ||:|
Db     211 GARFFNTFSFTMNIFGIGIKRALKGIVRSNVPEDMG-EHSIERISHLSEGYKDYMLTQVP 269

Qy     241 ALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNR------------VFI 287
           | |||:|:| ||:: ||:::|||| ||||:|  : |: :                :|:
Db     270 TLSKFAERYSDMVMKVLLSSLAGYVKAPWYKRWINRFKSLLTGEAYNPDEDIHLLKPIFV 329

Qy     288 PTKKFFNKEIREPSK-ALKEKVSTDTKDLFENKIGQGTVDFFN-------KEIRDPSKAL 339
           | : : |:  :| : |::| :     |    |        :|         : :| :
Db     330 DTPRNYIKDALKPLRDAVEENIVNPVSDYLRRKQNISRSQNYNDGHHKIDPSLYEPKRPH 389
```

```
Qy    340 KEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLFENKIGQGTVDFIN 399
          :| |:|  ::|: :          :| |:     |  :: : :    |  |
Db    390 IGIAANHARDYIDDKVNK----------AKELVSAAKDRATGIYADHVKPALSDITN 436

Qy    400 ---NEIRDPSKALIRKVY---TEADDLFENKIGQGTVDFINKEIRDPSKA 443
            |:: |     || :    :: |:  : |  :  |: :  |::    |
Db    437 VVKNDLLDAVN--IRNILRGSSQDDNNEQEKTEEEKVEEVKPELKQKEYA 484


RESULT 8
AAR25187
ID   AAR25187 standard; Protein; 91 AA.
XX
AC   AAR25187;
XX
DT   09-DEC-1992  (first entry)
XX
DE   21B4 gene clone product pT#13, EcoRI insert.
XX
KW   Beta-galactosidase; B. bovis; Bb; T21B4.
XX
OS   Babesia bovis.
XX
PN   EP492525-A.
XX
PD   01-JUL-1992.
XX
PF   20-DEC-1991;   91EP-0121990.
XX
PR   21-DEC-1990;   90AU-0004051.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Casu RE,  Commins MA;
XX
DR   WPI; 1992-218727/27.
DR   N-PSDB; AAQ26065.
XX
PT   Monoclonal antibody to Babesia bovis parasite - used to isolate
PT   antigens for use in vaccines for treating Babesiosis and
PT   providing immunity in cattle
XX
PS   Claim 14; Fig 7; 24pp; English.
XX
CC   The sequences given in AAR25186-89 are translation products of portions
CC   of the 21B4 gene which were isolated from a B. bovis (Bb) cDNA lambda
CC   gt11 library and cloned into pGEM7Zf(+). The resulting plasmids were
CC   tranformed into E. coli strain JM83.  The inserts were in frame,
CC   when translated, with the vector beta-galactosidase gene.  The fusion
CC   protiens produced by translation of these vectors were recognised by
CC   the monoclonal antibody of the invention, T21B4.  These fusion
CC   antigens could be used in vaccines for the treatment of babesiosis
CC   and to provide immunity in relation to Bb infection in cattle against
CC   different strains of Babesia by heterologous and homologous challenge.
XX
SQ   Sequence   91 AA;

  Query Match             6.8%;  Score 161;  DB 13;  Length 91;
  Best Local Similarity   35.2%;  Pred. No. 1e-05;
  Matches   32;  Conservative   15;  Mismatches   44;  Indels    0;  Gaps    0;

Qy      2 APSDSVGDVTKTLLAASESVDSAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCSA 61
          ||:: |||:  ||  |   :    : : |  ||   ||   :     :||  |: |||
Db      1 APAEVVGDLISTLETADTLMTLRDHMHNITKDMKHVLSNGREQIVNDVCSNAPEDSNCRE 60

Qy     62 SVSAYMSRCAKQDCLTLQSLKYPLEAKYQPL 92
          |: |  ||    | |: :::|||  :||||
Db     61 VVNNYADRCEMYGCFTIDNVRYPLYQEYQPL 91


RESULT 9
AAR25188
ID   AAR25188 standard; Protein; 56 AA.
XX
AC   AAR25188;
XX
DT   09-DEC-1992  (first entry)
XX
DE   21B4 gene clone product pT#13, EcoRI insert (2).
XX
KW   Beta-galactosidase; B. bovis; Bb; T21B4.
XX
OS   Babesia bovis.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 23
FT                   /label= Any_amino_acid
FT   Misc-difference 37
FT                   /label= Any_amino_acid
XX
PN   EP492525-A.
XX
PD   01-JUL-1992.
XX
PF   20-DEC-1991;   91EP-0121990.
XX
PR   21-DEC-1990;   90AU-0004051.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI   Casu RE,  Commins MA;
XX
DR   WPI; 1992-218727/27.
DR   N-PSDB; AAQ26066.
XX
PT   Monoclonal antibody to Babesia bovis parasite - used to isolate
PT   antigens for use in vaccines for treating Babesiosis and
PT   providing immunity in cattle
XX
PS   Claim 16; Fig 10; 24pp; English.
XX
CC   The sequences given in AAR25186-89 are translation products of portions
CC   of the 21B4 gene which were isolated from a B. bovis (Bb) cDNA lambda
CC   gt11 library and cloned into pGEM7Zf(+). The resulting plasmids were
CC   tranformed into E. coli strain JM83.  The inserts were in frame,
CC   when translated, with the vector beta-galactosidase gene.  The fusion
CC   protiens produced by translation of these vectors were recognised by
CC   the monoclonal antibody of the invention, T21B4.  These fusion
CC   antigens could be used in vaccines for the treatment of babesiosis
CC   and to provide immunity in relation to Bb infection in cattle against
CC   different strains of Babesia by heterologous and homologous challenge.
XX
SQ   Sequence   56 AA;

  Query Match             5.8%;  Score 138;  DB 13;  Length 56;
  Best Local Similarity   50.0%;  Pred. No. 0.0004;
  Matches   27;  Conservative    9;  Mismatches   18;  Indels    0;  Gaps    0;

Qy    161 SRYLYMATLYYKTYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLD 214
          :: |||||: ||||  |:   | |||: |||| :|   |:: |  ||| |:| |
Db      1 NKVLYMATMDYKTYLTVNSMNAXFFNRFSFTTKIFSXRIRQTLSDIIRWNVPED 54


RESULT 10
AAY93246
ID   AAY93246 standard; Protein; 974 AA.
XX
AC   AAY93246;
XX
DT   04-SEP-2000  (first entry)
XX
DE   An Escherichia coli virulence protein.
XX
KW   Virulence protein; tatA; tatB; tatC; tatE; mdoG; creC; recG; yggN;
KW   eck1; iroD; iroC; iroE; mtd2; ms1; vaccine; infection;
KW   Gram negative bacterium.
```

```
XX
OS   Escherichia coli.
XX
PN   WO200028038-A2.
XX
PD   18-MAY-2000.
XX
PF   09-NOV-1999;   99WO-GB03721.
XX
PR   09-NOV-1998;   98GB-0024569.
PR   09-NOV-1998;   98GB-0024570.
PR   17-DEC-1998;   98GB-0027814.
PR   17-DEC-1998;   98GB-0027815.
PR   17-DEC-1998;   98GB-0027816.
PR   17-DEC-1998;   98GB-0027818.
PR   13-JAN-1999;   99GB-0000708.
PR   13-JAN-1999;   99GB-0000710.
PR   13-JAN-1999;   99GB-0000711.
PR   28-JAN-1999;   99GB-0001915.
XX
PA   (MICR-) MICROSCIENCE LTD.
XX
PI   Crooke HR,  Clarke EE,  Everest PH,  Dougan G,  Holden DW,  Shea JE;
PI   Feldman RG;
XX
DR   WPI; 2000-376550/32.
DR   N-PSDB; AAA15186.
XX
PT   Peptide encoded by an operon including genes from Escherichia coli for
PT   screening potential drugs, detecting virulence and treating conditions
PT   associated with infection by a Gram negative bacterium -
XX
PS   Claim 2; Page 108-112; 122pp; English.
XX
CC   The present sequence represents an Escherichia coli virulence protein.
CC   The specification describes virulence proteins which are encoded
CC   by an operon including tatA, tatB, tatC, tatE, mdoG, creC, recG, yggN,
CC   eck1, iroD, iroC, iroE, mtd2 or ms1-16 genes obtained from Escherichia
CC   coli K1. The virulence proteins and polynucleotides, and their vaccines
CC   are useful for screening potential drugs, for the detection of virulence,
CC   and for treating or preventing conditions associated with infection by
CC   a Gram negative bacterium particularly Escherichia coli.
XX
SQ   Sequence   974 AA;

  Query Match             5.1%;  Score 121;  DB 21;  Length 974;
  Best Local Similarity   19.1%;  Pred. No. 0.45;
  Matches   89;  Conservative   77;  Mismatches  143;  Indels  156;  Gaps   22;

Qy      99 QLEAAFILFKE---SDANPANSTEKRFWMR----FRRGKNHSYFHDLVFNLLEKNVTRDA 151
           :|| | : |     ||:||      | ::       |     :|    ||| : :
Db     243 RLEKALLGFTNTMYSDSNPPIIARFRDYLEDGECIDRISESIFFTPQEFNLADHH----- 297

Qy     152 DATDIENFASRYLYMATLYYKTYTNVDEFGASFFNKLSFTTGLFG---------WGIKRA 202
             || :                 :|||     : | |  ::|         |
Db     298 ----IEGW----------------FNEFG-----QFSGTVSVYGEEPIHHVVTWKNNNQ 331

Qy     203 LKQ--------------IIRSNLPLDIGTEHSVSRLQHITSS------YKD------YMD 236
           | |              :  | ||:::      : |:  |         |:|      | |
Db     332 LTQCGPFKIKLAYIHGRLRDSRLPMELW-----APLKEKTDRYGGLYIYRDGLRILPYGD 386

Qy     237 TQIPALPKFAKRFSLMV------VQRLLATV---------------AGYVDTPWYKKWY 274
           :    | |  || :|         :||   :                 |:::   ||::
Db     387 SDTDFL-KIEKRRTLSASEYFFSYRRLFGAIELTKENNASLVEKAGREGFIENKPYKQFK 445

Qy     275 MKLKNFMV-------------NRVFIPTKKFFNKE---IREPSKALKEKVSTDTKDLFEN 318
             |:|| :               : :|: ||:  |:|   : : ||  | |     |||:
Db     446 EMLENFFIEIARDFFKDDGDMSELFVETKQRRNEEHDLLSKRSKQTKAKKDRLKKDLY-- 503

Qy     319 KIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFEN-KIGQGTVDFINNEIRDPSKALIRK 377
                  |||:|   |       |: |  :: | : :|    :|:: |:|:: : |:|:
Db     504 -------DFFDKLDNDYWNIEINKLINKNEEYFSSTEITDTNIDYVYNKIKEQNDAIIKN 556

Qy     378 VSTGAEDLFENKIGQGTVDFINNEIRDPSKALIRKVYTEADDLFEN------KIGQGTVD 431
           :    :   :   | |    ::| : |  :   :|:    ::| :|      ::     |
Db     557 LRNSVD--IKKPSGVGLTKELSN-LWDRYQIERQKILLSLNELKDNVDRKLIELDNKNND 613

Qy     432 FIN--KEIRD-----------------PSKALIRKVSTEADNLL 456
           |:|  | : |                  :|  :: | ::|: |:
Db     614 FLNLRKRLEDSLNLQQSYYEKELTKLYNDAKNALKDVQSKANRLI 658


RESULT 11
AAG67416
ID   AAG67416 standard; Protein; 1038 AA.
XX
AC   AAG67416;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Amino acid sequence of bimC homologue, cin8.
XX
KW   bimC; kinesin related protein; fungal viability; antifungal; cin8;
KW   fungal infection.
XX
OS   Saccharomyces cerevisiae.
XX
PN   US6284480-B1.
XX
PD   04-SEP-2001.
XX
PF   03-APR-2000; 2000US-0541782.
XX
PR   03-APR-2000; 2000US-0541782.
XX
PA   (CYTO-) CYTOKINETICS INC.
XX
PI   Nislow CE,  Sakowicz R,  Beraud C;
XX
DR   WPI; 2001-540724/60.
DR   N-PSDB; AAH78010.
XX
PT   Identifying a modulator, e.g. antifungal agent, of a target protein
PT   comprising bimC or its fragment by determining enzymatic activity of a
PT   reaction, in the presence and absence of the compound, that uses ADP or
PT   phosphate produced by bimC -
XX
PS   Disclosure; Fig 4; 47pp; English.
XX
CC   The present sequence represents bimC homologue, designated cin8. BimC is
CC   a kinesin related protein, which is essential for fungal viability. The
CC   specification describes a method of identifying modulators of bimC.
CC   The method comprises adding a test agent to a mixture comprising bimC
CC   protein that directly or indirectly produces ADP or phosphate, subjecting
CC   the mixture to an enzymatic reaction that uses the ADP or phosphate,
CC   and determining the enzymatic activity in presence and absence of test
CC   compound. A change in the activity level between the presence and absence
CC   of the candidate agent indicates a modulator of the target protein
CC   function. The method is useful for identifying a modulator, e.g.
CC   antifungal agents, of bimC. The modulators can be used, for example, to
CC   inhibit the growth or spread of fungi, mould, fruit flies, etc.. The
CC   modulators can be used for preventing and treating infections caused
CC   by Chytridiomycetes, Hyphochrytridiomycetes, Plasmodiophoromycetes,
CC   Oomycetes, Zygomycetes, Ascomycetes, and Basidiomycetes.
XX
SQ   Sequence   1038 AA;

  Query Match             5.1%;  Score 120;  DB 22;  Length 1038;
  Best Local Similarity   19.5%;  Pred. No. 0.6;
  Matches  119;  Conservative   97;  Mismatches  214;  Indels  180;  Gaps   28;

Qy       4 SDSVGDVTKTLLAASESVDSAANAYMI--------NSDMSDYLSAVSD---------NFA 46
           ||: | : : ||   ::::   | |::         | :: | | : |:          |
Db     188 SDAAGIIPRVLLKLFDTLELQQNDYVVKCSFIELYNEELKDLLDSNSNGSSNTGFDGQFM 247

Qy      47 ERICSQVPKGSNCSASVSAYMSRCAKQDC-------LTLQS--LKYPLEAKYQPLTLPDP 97
```

```
          :::       :| : | ||  ||   ::        || ::  |:  |  |  | |:
Db    248 KKLRIFASSTANNTTSNSASSSRSNSRNSSPRSLNDLTPKAALLRKRLRTKSLPNTIKQQ 307

Qy     98 YQLEAAFILFKESDANPANSTE-----------KRFWMRFRRGKNHSYFHDLV------- 139
          || : |      | :|  ::|           :|  |      |  |  :|
Db    308 YQQQQAVNSRNNSSSNSGSTTNNASSNTNTNNGQRSSMAPNDQTNGIYIQNLQEFHITNA 367

Qy    140 ---FNLLEKNVT-RDADATDIENFASR--YLYMATLYYKTY----------------TN 176
              |||:| :  |   :| : :|:||   ::  ||| |                    |
Db    368 MEGLNLLQKGLKHRQVASTKMNDFSSRSHTIFTITLYKKHQDELFRISKMNLVDLAGSEN 427

Qy    177 VDEFG--------ASFFNKLSFTTG-------------------------LFGWGIKRA 202
          ::  |        |   |:   | |                          |  | |
Db    428 INRSGALNQRAKEAGSINQSLLTLGRVINALVDKSGHIPFRESKLTRLLQDSLGGNTKTA 487

Qy    203 LKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALPKFAKRFSLM---------V 253
          |   | :  |  : :| : | |::  |  |:  :   | |  |  :  |:         :
Db    488 L---IATISPAKVTSEETCSTLEY-ASKAKNIKNK--PQLGSFIMKDILVKNITMELAKI 541

Qy    254 VQRLLATVAG---YVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREPSKALKEKVST 310
             ||:| :    |:    ||     |::  | |          :| :   ::|  | :
Db    542 KSDLLSTKSKEGIYMSQDHYKNLNSDLESYK-NEV---------QECKREIESLTSKNAL 591

Qy    311 DTKDLFENKIGQGTVDFFNKEIR------DPSKALKEKVSNDAKDL--FENKIGQGT--- 359
            ||  ::|    |:   | :|       | :|  :|      ::  | ||: : |
Db    592 LVKDKLKSK---ETIQSQNCQIESLKTTIDHLRAQLDKQHKTEIEISDFNNKLQKLTEVM 648

Qy    360 -----------VDF-------INNEIRDPSKALIRKVSTGAED--LFENKIGQGTVDFIN 399
                    :|      |  ||:     |  :::|  ::  | |  | |  :| |
Db    649 QMALHDYKKRELDLNQKFEMHITKEIKKLKSTLFLQLNTMQQESILQETNI-QPNLDMIK 707

Qy    400 NEIRDPSKAL-----------IRKVYTEADDLFENKIGQGTVDFINKEIRDPSKAL---I 445
          ||:    : :           ::|:  |:   |   | :  :| |  : :   | :   :
Db    708 NEVLTLMRTMQEKAELMYKDCVKKILNESPKFFNVVIEK--IDIIRVDFQKFYKNIAENL 765

Qy    446 RKVSTEADNL 455
            :| | :|:
Db    766 SDISEENNNM 775


RESULT 12
AAR46608
ID   AAR46608 standard; Protein; 1663 AA.
XX
AC   AAR46608;
XX
DT   22-SEP-1994  (first entry)
XX
DE   Plasmodium falciparum erythrocyte membrane protein PfEMP3.
XX
KW   Plasmodium falciparum erythrocyte membrane protein; PfEMP3;
KW   malaria; antigen; epitope; vaccine; anti-idiotype antibody.
XX
OS   Plasmodium falciparum (Malayan Camp strain).
XX
FH   Key             Location/Qualifiers
FT   Region          472..493
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          494..515
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          516..537
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          538..559
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          560..581
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          582..603
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          604..625
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          626..647
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          648..669
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          670..691
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          692..713
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          714..735
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          736..757
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          758..779
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          780..801
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          802..823
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          824..845
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          846..867
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          868..889
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          890..911
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          912..933
FT                   /label= tandem_repeat
FT                   /note= "one of 21 complete segments of homology
FT                   of 22 amino acid length"
FT   Region          934..946
FT                   /label= partial_tandem_repeat
FT   Region          949..967
FT                   /label= tandem_repeat
FT                   /note= "one of 11 complete segments of homology
FT                   of 19 amino acid length"
```

```
FT   Region          968..986
FT                   /label= tandem_repeat
FT                   /note= "one of 11 complete segments of homology
FT                          of 19 amino acid length"
FT   Region          987..1005
FT                   /label= tandem_repeat
FT                   /note= "one of 11 complete segments of homology
FT                          of 19 amino acid length"
FT   Region          1006..1024
FT                   /label= tandem_repeat
FT                   /note= "one of 11 complete segments of homology
FT                          of 19 amino acid length"
FT   Region          1025..1043
FT                   /label= tandem_repeat
FT                   /note= "one of 11 complete segments of homology
FT                          of 19 amino acid length"
FT   Region          1044..1062
FT                   /label= tandem_repeat
FT                   /note= "one of 11 complete segments of homology
FT                          of 19 amino acid length"
FT   Region          1063..1081
FT                   /label= tandem_repeat
FT                   /note= "one of 11 complete segments of homology
FT                          of 19 amino acid length"
FT   Region          1082..1100
FT                   /label= tandem_repeat
FT                   /note= "one of 11 complete segments of homology
FT                          of 19 amino acid length"
FT   Region          1101..1119
FT                   /label= tandem_repeat
FT                   /note= "one of 11 complete segments of homology
FT                          of 19 amino acid length"
FT   Region          1120..1138
FT                   /label= tandem_repeat
FT                   /note= "one of 11 complete segments of homology
FT                          of 19 amino acid length"
FT   Region          1139..1157
FT                   /label= tandem_repeat
FT                   /note= "one of 11 complete segments of homology
FT                          of 19 amino acid length"
FT   Region          1158..1173
FT                   /label= partial_tandem_repeat
FT                   /note= "partial segment of homology"
FT   Region          1179..1193
FT                   /label= tandem_repeat
FT                   /note= "one of 4 complete segments of homology
FT                          of 15 amino acid length"
FT   Region          1194..1208
FT                   /label= tandem_repeat
FT                   /note= "one of 4 complete segments of homology
FT                          of 15 amino acid length"
FT   Region          1209..1223
FT                   /label= tandem_repeat
FT                   /note= "one of 4 complete segments of homology
FT                          of 15 amino acid length"
FT   Region          1224..1238
FT                   /label= tandem_repeat
FT                   /note= "one of 4 complete segments of homology
FT                          of 15 amino acid length"
FT   Region          1248..1260
FT                   /label= tandem_repeat
FT                   /note= "one of 27 complete segments of homology
FT                          of 13 amino acid length"
FT   Region          1261..1273
FT                   /label= tandem_repeat
FT                   /note= "one of 27 complete segments of homology
FT                          of 13 amino acid length"
FT   Region          1274..1286
FT                   /label= tandem_repeat
FT                   /note= "one of 27 complete segments of homology
FT                          of 13 amino acid length"
FT   Region          1287..1299
FT                   /label= tandem_repeat
FT                   /note= "one of 27 complete segments of homology
FT                          of 13 amino acid length"
FT   Region          1300..1312
FT                   /label= tandem_repeat
FT                   /note= "one of 27 complete segments of homology
FT                          of 13 amino acid length"
FT   Region          1313..1325
FT                   /label= tandem_repeat
FT                   /note= "one of 27 complete segments of homology
FT                          of 13 amino acid length"
FT   Region          1326..1338
FT                   /label= tandem_repeat
FT                   /note= "one of 27 complete segments of homology
FT                          of 13 amino acid length"
FT   Region          1339..1351
FT                   /label= tandem_repeat
FT                   /note= "one of 27 complete segments of homology
FT                          of 13 amino acid length"
FT   Region          1352..1364
FT                   /label= tandem_repeat
FT                   /note= "one of 27 complete segments of homology
FT                          of 13 amino acid length"
FT   Region          1365..1377

  Query Match             5.1%;  Score 119.5;  DB 15;  Length 1663;
  Best Local Similarity   31.4%;  Pred. No. 1.2;
  Matches   58;  Conservative   31;  Mismatches   65;  Indels   31;  Gaps   14;

Qy     294 NKEIR-EPSKALKEKVSTDTKDLFENKIGQG---TVDFFNKEIRDP-SKALKEKVSNDAK 348
           |||:| : |: |||      |:| :||  :|     :  |||:::  |: |||      |
Db    1045 NKELRNKGSEGLKENAELKNKEL-QNKGSEGLKENAELKNKELQNKGSEGLKENAEQKNK 1103

Qy     349 DLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLFE-------NKIGQG---TVDFI 398
           :| :||  :|  :  | |::  :| |  | | | ::  |       ||   |     :
Db    1104 EL-QNKGSEGLKE--NAELK--NKELRNKGSDGLKENAELKNKELRNKGSDGLKENAELK 1158

Qy     399 NNEIRDP-SKALIRKVYTEADDLFENKIGQGTVDFINKEIRDPSKALIRKVSTEAD---- 453
           | |:|:  |: |   |||  :||  | |     |  ||:::  :| |: |  :  |
Db    1159 NKELRNKGSEGLKENVYTN-NDLKNNDI--QNKDLSNKDMK--NKELLNKDISNKDMKNK 1213

Qy     454 NLLEK 458
           || |
Db    1214 ELLNK 1218


RESULT 13
AAR46605
ID   AAR46605 standard; Protein; 1588 AA.
XX
AC   AAR46605;
XX
DT   22-SEP-1994  (first entry)
XX
DE   Malarial PfEMP3 epitopic fragment.
XX
KW   Plasmodium falciparum erythrocyte membrane protein; PfEMP3;
KW   malaria; antigen; epitope; vaccine; anti-idiotype antibody.
XX
OS   Plasmodium falciparum (Malayan Camp strain).
XX
PN   WO9403604-A.
XX
PD   17-FEB-1994.
XX
PF   05-AUG-1993;   93WO-US07261.
XX
PR   07-AUG-1992;   92US-0927531.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Handunnetti SM,  Howard RJ,  Pasloske BL,  Van Schravendijk MR;
XX
DR   WPI; 1994-065693/08.
```

```
DR   N-PSDB; AAQ70102.
XX
PT   New malaria antigen, PfEMP3 - used to isolate and produce prods.
PT   for use in diagnosis, therapy and prevention of malarial
PT   infection
XX
PS   Claim 12; Page 79-85; 79pp; English.
XX
CC   The PfEMP3 malarial antigen is recognised by monoclonal antibody MAb
CC   12C11. Nucleic acid sequences encoding part of the 315kD antigen,
CC   have been isolated and sequenced. PfEMP3 is encoded on chromosome 2
CC   of the P.falciparum genome and is thought to be associated with knob
CC   formation and structure; malarial strains carrying deletions of the
CC   gene coding for PfEMP3 exhibit a knobless phenotype.
XX
SQ   Sequence   1588 AA;

  Query Match             5.0%;  Score 118.5;  DB 15;  Length 1588;
  Best Local Similarity   30.8%;  Pred. No. 1.4;
  Matches   57;  Conservative   31;  Mismatches   66;  Indels   31;  Gaps   14;

Qy     294 NKEIR-EPSKALKEKVSTDTKDLFENKIGQG---TVDFFNKEIRDP-SKALKEKVSNDAK 348
           |||:| : |: |||      |:| :||  :|     :  |||:::  |: |||      |
Db    1045 NKELRNKGSEGLKENAELKNKEL-QNKGSEGLKENAELKNKELQNKGSEGLKENAEQKNK 1103

Qy     349 DLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLFE-------NKIGQG---TVDFI 398
           :| :||  :|  :  | |::  :| |  | | | ::  |       ||   |     :
Db    1104 EL-QNKGSEGLKE--NAELK--NKELRNKGSDGLKENAELKNKELRNKGSDGLKENAELK 1158

Qy     399 NNEIRDP-SKALIRKVYTEADDLFENKIGQGTVDFINKEIRDPSKALIRKVSTEAD---- 453
           | |:|:  |: |   |||  |   :|   |   |  ||:::  :| |: |  :  |
Db    1159 NKELRNKGSEGLKENVYTNND--LKNNDNQNK-DLSNKDMK--NKELLNKDISNKDMKNK 1213

Qy     454 NLLEK 458
           || |
Db    1214 ELLNK 1218


RESULT 14
AAB18161
ID   AAB18161 standard; Protein; 2441 AA.
XX
AC   AAB18161;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related protein SEQ ID NO:18.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 50-57; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence   2441 AA;

  Query Match             5.0%;  Score 118.5;  DB 21;  Length 2441;
  Best Local Similarity   31.4%;  Pred. No. 2.5;
  Matches   58;  Conservative   32;  Mismatches   64;  Indels   31;  Gaps   14;

Qy     294 NKEIR-EPSKALKEKVSTDTKDLFENKIGQG---TVDFFNKEIRDP-SKALKEKVSNDAK 348
           |||:: : |: |||      |:| :||  :|     :  |||:|:  |: |||      |
Db    1121 NKELQNKGSEGLKENAELKNKEL-QNKGSEGLKENAELKNKELRNKGSEGLKENAELKNK 1179

Qy     349 DLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLFE-------NKIGQG---TVDFI 398
           :| :||  :|  :  | |::  :| |  | | | ::  |       ||  :|     :
Db    1180 EL-QNKGSEGLKE--NAELK--NKELQNKGSEGLKENAELKNKELQNKGSEGLKENAELK 1234

Qy     399 NNEIRDP-SKALIRKVYTEADDLFENKIGQGTVDFINKEIRDPSKALIRKVSTEAD---- 453
           | |:|:  |: |   |||  :||  | |     |  ||:::  :| |: |  :  |
Db    1235 NKELRNKGSEGLKENVYTN-NDLKNNDI--QNKDLSNKDMK--NKELLNKDISNKDMKNK 1289

Qy     454 NLLEK 458
           || |
Db    1290 ELLNK 1294


RESULT 15
ABB59227
ID   ABB59227 standard; Protein; 1480 AA.
XX
AC   ABB59227;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 4473.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
```

```
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03330.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 4473; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1480 AA;

  Query Match             4.8%;  Score 113.5;  DB 22;  Length 1480;
  Best Local Similarity   21.3%;  Pred. No. 3.3;
  Matches   93;  Conservative   52;  Mismatches  141;  Indels  151;  Gaps   20;

Qy       3 PSDSVGDVTKTLLAASESV---------DSAANAYMINSDMSDYL----SAVSDNFAERI 49
           |   |||    :|  :| :         : |    :: | ::: |    ||::|:| : :
Db    1108 PDAVVGDSDAKVLTLAEHIIAAMPERIKNEADEMMVLGSPLTEPLTSESSALTDSFMDPL 1167

Qy      50 CSQVPK--------------------GSNCS------------------------ASV 63
             :|                      : ||                        :|
Db    1168 LDSLPNTHFDSDFSFDFHDHSYRYHDVSTPCSSLSPASSGPLQSPASYSILGTDPSVSSP 1227

Qy      64 SAYMSRCAKQDCLTLQSL-KYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANSTEKRF 122
           |   |     : |   |:  || || :  ||| |  | |    |::      |   :| :
Db    1228 SPPPSTKQLTEFLHASSISSYPFEADFSKLTLTDTEQRE----LYE-----AAKCIQKAY 1278

Qy     123 WMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEFGA 182
             :|:           | |:|  | | || |:|:  ||   |  ||:  ||
Db    1279 --RSYKGRQ---------KLEEQNKERSA-ATVIQNYYRRYKQYA--YYRQMTN------ 1318

Qy     183 SFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEH-SVSRLQHITSSYKDYMDTQIPA 241
               :      |:   |  |:  :| | |   ::| |||      ||: |:
Db    1319 ------AALVIQHGYRSYRRNKRFKKSGLCLSSSSDHGSVSSNSQCLSSFYDHYKQD--- 1369

Qy     242 LPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREPS 301
                    |: |  :     ||                          ||   ||
Db    1370 ------------QQQLHELGSQPSTP-------------------------KE-TSPS 1389

Qy     302 KALKEKVSTDTKDLFENKIGQGTVDFFNKEIR---DPSKALKEKVSNDAK-DLFENKIGQ 357
             ||   |  |::    || |     | :: |      :| |||: :  : :  :|   |
Db    1390 GPLKRTYSQSTQNQAARKIQQ-----FMRQSRIKLQKERAEKEKLVHQRRAEYLQNLQFQ 1444

Qy     358 GTVDFI---NNEIRDPS 371
           |  : :    | |  ||
Db    1445 GQQEMLVYHENNISAPS 1461


RESULT 16
AAY02371
ID   AAY02371 standard; Protein; 474 AA.
XX
AC   AAY02371;
XX
DT   13-JUL-1999  (first entry)
XX
DE   Polypeptide identified by the signal sequence trap method.
XX
KW   Signal sequence trap method; SST method; immunisation; inhibition;
KW   infection; allergy; cancer; regulation; tissue formation; tissue repair;
KW   activin activity; inhibin activity; chemokine activity;
KW   cytokine activity; blood coagulation regulation; agonist; antagonist;
KW   metabolic disorder; hormonal disorder; immune disorder; wound;
KW   severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer.
XX
OS   Homo sapiens.
XX
PN   WO9918126-A1.
XX
PD   15-APR-1999.
XX
PF   06-OCT-1998;   98WO-JP04514.
XX
PR   07-OCT-1997;   97JP-0274674.
XX
PA   (ONOY ) ONO PHARM CO LTD.
XX
PI   Fukushima D,  Shibayama S,  Tada H;
XX
DR   WPI; 1999-277254/23.
DR   N-PSDB; AAX35720, AAX35721.
XX
PT   Polypeptides identified by the signal sequence trap method from a
PT   human cDNA library
XX
PS   Claim 1; Page 137-139; 281pp; Japanese.
XX
CC   AAY02358-84 represent novel polypeptides which are identified from a
CC   human placental cDNA library by the signal sequence trap (SST) method.
CC   The polypeptides are encoded by the cDNA seqeuences in AAX35694-X35747.
CC   The polypeptides have a broad range of physiological activity, including
CC   immunisation against and inhibition of infections, allergies and cancer;
CC   regulation of tissue formation and repair; activin/inhibin activity;
CC   chemokine/cytokine activity; blood coagulation regulation; and
CC   receptor/ligand agonist or antagonist activity. The polypeptides can
CC   be used for prevention and treatment of disorders including infections
CC   by bacteria, yeasts and viruses (including HIV) and protozoa; metabolic
CC   and hormonal disorders; immune disorders (including severe combined
CC   immunodeficiency (SCID) and AIDS; thrombosis; cancer; and traumatic or
CC   surgical wounds.
XX
SQ   Sequence   474 AA;

  Query Match             4.7%;  Score 111;  DB 20;  Length 474;
  Best Local Similarity   20.4%;  Pred. No. 1.1;
  Matches   95;  Conservative   79;  Mismatches  165;  Indels  126;  Gaps   26;

Qy      62 SVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLE------------AAFILFK 108
           ::  | |  :  |  :   :   |::| : |   | ::||            |     :
Db      61 TIKCYESFMSGAD--SFDEMNAELQSKLKDLFNVDAFKLESLEAKNRALNEQIARLEQER 118

Qy     109 ESDANPANSTEK----------RFWMRFRRGKNHSYFHDLVFNLLEKNVTR--------D 150
           | : |   |  |            ::       ::||   |   | | : : |
Db     119 EKEPNRLESLRKLKASLQGDVQKYQAYMSNLESHSAILDQKLNGLNEEIARVELECETIK 178

Qy     151 ADATDIENFASRYLYMATLYYKTYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSN 210
           : | ::|      |       :  :::                        :  |:| | :
Db     179 QENTRLQNIIDNQKY-------SVADIERINHE---------------RNELQQTI-NK 214

Qy     211 LPLDIGTEHSVSRLQHITSSY---KDYMDTQIPALPKFAKRFSLMVVQRLLATVAGYVDT 267
           |  |:  |    :| :    |   |: ::||:    | |::  | :: :      ||
Db     215 LTKDLEAEQ--QKLWNEELKYARGKEAIETQLAEYHKLARK--LKLIPKGAENSKGY--- 267

Qy     268 PWYKKWYMKLK-----NFMVN---RVFIPTKKFFNKEIREPSKALKEKVS-TDTKD---- 314
           : :|      | :|    :|::| |:  |:   | :||| :|:   || :
Db     268 ----DFEIKFNPEAGANCLVKYRAQVYVPLKELLNETEEEINKALNKKMGLEDTLEQLNA 323

Qy     315 -LFENKIGQGTV--------DFFNKEIRDPSKALKEKVSNDA------KDLFENKIGQGT 359
           : |:|   ||:        | : ::|::  :   || :::       | | |: : ||
Db     324 MITESKRSVGTLKEEVQKLDDLYQQKIKEAEEE-DEKCASELESLEKHKHLLESTVNQGL 382
```

```
Qy    360 VDFINNEIRDPSKALIRKVSTGAEDLFENKIG---QGTVDFINNEIRDPSKAL---IRKV 413
          : : ||:    :     | |  |:    |:|   |  :: :   :    | |   | ||
Db    383 SEAM-NELDAVQREYQLVVQTTTEE--RRKVGNNLQRLLEMVATHVGSVEKHLEEQIAKV 439

Qy    414 ---YTE--ADDLFENKIGQGTVDFINKEIRD--PSKALIRKVSTE 451
             | |  ::|| ||           ||||    || : | | |
Db    440 DREYEECMSEDLSEN----------IKEIRDKYEKKATLIKSSEE 474


RESULT 17
AAW22230
ID   AAW22230 standard; Protein; 885 AA.
XX
AC   AAW22230;
XX
DT   12-SEP-1997  (first entry)
XX
DE   K. lactis origin of replication complex protein 1.
XX
KW   Origin of replication complex; ORC; yeast; Kluyveromyces lactis;
KW   chromatography; peptide sequencing; primer; amplification; PCR; genome;
KW   polymerase chain reaction; open reading frame; cell growth; cancer;
KW   infection; inflammation; hypersensitivity.
XX
OS   Kluyveromyces lactis.
XX
PN   US5614618-A.
XX
PD   25-MAR-1997.
XX
PF   16-DEC-1993;   93US-0168479.
XX
PR   07-JUN-1995;   95US-0484106.
PR   16-DEC-1993;   93US-0168479.
XX
PA   (COLD-) COLD SPRING HARBOR LAB.
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Bell SP,  Foss M,  Gavin K,  Herskowitz I,  Hidaka M;
PI   Kobayashi R,  Laurenson P,  Li J,  McNally FJ,  Rine J;
PI   Stillman BW;
XX
DR   WPI; 1997-201534/18.
DR   N-PSDB; AAT73285.
XX
PT   Nucleic acids encoding origin of replication complex proteins - used
PT   for screening for lead cpds. for therapy or diagnosis of disease
PT   associated with undesirable cell growth
XX
PS   Claim 1; Column 61-66; 53pp; English.
XX
CC   This is the amino acid sequence of the origin of replication complex
CC   protein 1 (ORC1) from the yeast Kluyveromyces lactis.  The sequence was
CC   isolated using primers based on amino acid sequence conserved between
CC   the ORC1 and SIR3 proteins from Saccharomyces cerevisiae.  The amplified
CC   fragment was then used for low stringency DNA hybridisation to obtain
CC   the K. lactis ORC1 gene sequence.  The ORC proteins (AAW22224-35) can be
CC   used to screen chemical libraries to identify lead compounds useful in
CC   treatment and diagnosis of undesired cell growth, e.g. cancer,
CC   infections, inflammation and hypersensitivity.
XX
SQ   Sequence   885 AA;

  Query Match             4.7%;  Score 111;  DB 18;  Length 885;
  Best Local Similarity   19.3%;  Pred. No. 2.6;
  Matches   98;  Conservative   82;  Mismatches  199;  Indels  128;  Gaps   22;

Qy      5 DSVGDVTKTLLAASESVDSAANAYMINSD--MSDYLSAVSDNFAERICSQV--------- 53
          ::: |    |    || :  |||   :||    || ||      |    :|
Db    274 EAISDNESDLSEYHESKEEFANASSSDSDEEFEDYQSAEELAIVEPAKKKVRSIKPDIPI 333

Qy     54 -PKGSNCSASVSAYMS---RCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKE 109
           |  |      ||  |   :  | : :  :    |   :|:   :||       |  :

Db    334 SPVKSQTPLQPSAVHSSPRKFFKNNIVRAKKAYTPFSKRYKNPKIPDLND-----IFQRH 388

Qy    110 SDANPANSTEKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRD-ADATDIENF-ASRYLYMA 167
          ::     : |:||     :||  : |  :    |  :|   :    | | :|:  :|    |
Db    389 NNDLDIAALEERFRTVSAKGKMETIFSKVKKQLNSRNSKEEIVKAADFDNYLPARENEFA 448

Qy    168 TLYYKTYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGT---EHSVSRL 224
          ::|   |: : | | |    :: | |:   |    ::::::     |: |   :  : |
Db    449 SIYLSLYSAI-EAGTSTSIYIAGTPGV---GKTLTVREVVK-----DLMTSADQKELPRF 499

Qy    225 QHI----------TSSYK----------------DYMDTQIPALPKFAKRFSLMVVQRL 257
          |:|          : ||:                  : ::     :|   ||  :::: |
Db    500 QYIEINGLKIVKASDSYEVFWQKISGEKLTSGAAMESLEFYFNKVPATKKRPIVVLLDEL 559

Qy    258 LATVAGYVDTPW-YKKW--YMKLKNFMV---NRVFIPTKKFFNK---------------- 295
           | |:   |  : :  |  |   |  :|   | : :| :   ||
Db    560 DALVSKSQDVMYNFFNWATYSNAKLIVVAVANTLDLPERHLGNKISSRIGFTRIMFTGYT 619

Qy    296 --EIR----------------EPSKALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPS 336
             |:|                 :|     :| |:  : |    :   || |
Db    620 HEELRTIINLRLKYLNESSFYVDPETGSSYMISPDSSTI-ETDEEEKRKDFSN------Y 672

Qy    337 KALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDL-----FENKI- 390
          | || ::: || ::   ||       :: ::|   | : | |     |      :|  :
Db    673 KRLKLRINPDAIEIASRKIAS-----VSGDVRRALKVVKRAVEYAENDYLKRLRYERLVN 727

Qy    391 --------GQGTVDFINNEIRDPSKAL 409
                 | |  :  : ||:  :|||
Db    728 SKKDTSGNGTGNEELQSVEIKHITKAL 754


RESULT 18
AAW14136
ID   AAW14136 standard; Protein; 885 AA.
XX
AC   AAW14136;
XX
DT   23-JUL-1997  (first entry)
XX
DE   Kluyveromyces lactis origin of replication complex ORC1.
XX
KW   Origin of replication complex; ORC; gene therapy; cancer;
KW   neoplasia; inflammation; hypersensitivity.
XX
OS   Kluyveromyces lactis.
XX
PN   WO9640977-A1.
XX
PD   19-DEC-1996.
XX
PF   07-JUN-1996;   96WO-US09403.
XX
PR   07-JUN-1995;   95US-0484105.
XX
PA   (COLD-) COLD SPRING HARBOR LAB.
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Bell SP,  Foss M,  Herskowitz I,  Kobayashi R,  Laurenson P;
PI   Li JJ,  McNally FJ,  Rine J,  Stillman BW;
XX
DR   WPI; 1997-052354/05.
DR   N-PSDB; AAT62358.
XX
PT   Nucleic acid encoding origin of replication complex (ORC) protein -
PT   useful to screen for lead pharmaceuticals capable of disrupting ORC
PT   protein function, and inhibiting cell growth
XX
PS   Disclosure; Page 18-22; 57pp; English.
XX
CC   Origin of replication (ORC) proteins (AAW14136-41) are respectively
CC   encoded by cDNA clones (AAT62358-63) from Kluyveromyces lactis,
CC   Schizosaccharomyces pombe, human (ORC1), Arabidopsis thaliana,
CC   Caenorhabditis elegans and human (ORC2).  The ORC polypeptides
```

```
CC   can be produced in transformed host cells, and in transgenic
CC   animals for functional studies (e.g. the efficacy of candidate
CC   drugs for diseases associated with expression of ORC).  The
CC   recombinant ORC proteins can be used in a novel method of screening
CC   for lead pharmaceuticals, esp. cpds. capable of disrupting ORC
CC   function and inhibiting cell growth, useful in the treatment of
CC   neoplasia, inflammation, hypersensitivity, etc.
XX
SQ   Sequence   885 AA;

  Query Match             4.7%;  Score 111;  DB 18;  Length 885;
  Best Local Similarity   19.3%;  Pred. No. 2.6;
  Matches   98;  Conservative   82;  Mismatches  199;  Indels  128;  Gaps   22;

Qy       5 DSVGDVTKTLLAASESVDSAANAYMINSD--MSDYLSAVSDNFAERICSQV--------- 53
Db     274 EAISDNESDLSEYHESKEEFANASSSDSDEEFEDYQSAEELAIVEPAKKKVRSIKPDIPI 333

Qy      54 -PKGSNCSASVSAYMS---RCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKE 109
Db     334 SPVKSQTPLQPSAVHSSPRKFFKNNIVRAKKAYTPFSKRYKNPKIPDLND-----IFQRH 388

Qy     110 SDANPANSTEKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRD-ADATDIENF-ASRYLYMA 167
Db     389 NNDLDIAALEERFRTVSAKGKMETIFSKVKKQLNSRNSKEEIVKAADFDNYLPARENEFA 448

Qy     168 TLYYKTYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGT---EHSVSRL 224
Db     449 SIYLSLYSAI-EAGTSTSIYIAGTPGV---GKTLTVREVVK-----DLMTSADQKELPRF 499

Qy     225 QHI----------TSSYK----------------DYMDTQIPALPKFAKRFSLMVVQRL 257
Db     500 QYIEINGLKIVKASDSYEVFWQKISGEKLTSGAAMESLEFYFNKVPATKKRPIVVLLDEL 559

Qy     258 LATVAGYVDTPW-YKKW--YMKLKNFMV---NRVFIPTKKFFNK---------------- 295
Db     560 DALVSKSQDVMYNFFNWATYSNAKLIVVAVANTLDLPERHLGNKISSRIGFTRIMFTGYT 619

Qy     296 --EIR----------------EPSKALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPS 336
Db     620 HEELRTIINLRLKYLNESSFYVDPETGSSYMISPDSSTI-ETDEEEKRKDFSN------Y 672

Qy     337 KALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDL-----FENKI- 390
Db     673 KRLKLRINPDAIEIASRKIAS-----VSGDVRRALKVVKRAVEYAENDYLKRLRYERLVN 727

Qy     391 --------GQGTVDFINNEIRDPSKAL 409
Db     728 SKKDTSGNGTGNEELQSVEIKHITKAL 754


RESULT 19
AAR08259
ID   AAR08259 standard; protein; 566 AA.
XX
AC   AAR08259;
XX
DT   05-MAR-1991  (first entry)
XX
DE   Haemagglutinin.
XX
KW   Influenza; HA; ribosomal frameshift signal sequence;
KW   membrane anchor; RFS; ss.
XX
OS   Influenza virus A/PR8/34.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..17
FT                   /label= signal peptide
FT   Protein         18..344
FT                   /label= HA1
FT   Protein         345..564
FT                   /label= HA2
FT   Region          534..551
FT                   /label= anchor region substd. with RFS
FT   Modified-site   27..30
FT                   /label= N-glycosylation site
FT   Modified-site   40..42
FT                   /label= N-glycosylation site
FT   Modified-site   144..146
FT                   /label= N-glycosyaltion site
FT   Modified-site   304..306
FT                   /label= N-glycosylation site
FT   Modified-site   498..500
FT                   /label= N-glycosylation site
FT   Modified-site   557..559
FT                   /label= N-glycosylation site
XX
PN   WO9014422-A.
XX
PD   29-NOV-1990.
XX
PF   21-MAY-1990;   90WO-GB00791.
XX
PR   19-MAY-1989;   89GB-0011555.
XX
PA   (LYNX-) LYNXVALE LTD.
XX
PI   Inglis SC,  Brierley I;
XX
DR   WPI; 1990-375989/50.
XX
PT   Ribosomal frame shifting signal sequences - isolated from
PT   infectious bronchitis virus genomic RNA and used in protein
PT   prodn.
XX
PS   Disclosure; Fig 19; 55pp; English.
XX
CC   The HA gene encodes a spike-like protein which is embedded in the
CC   membrane via a hydrophobic anchor sequence.  A portion of this
CC   anchor sequence may be replaced with a ribosomal frame shift signal
CC   sequence (RFS), in such a way that ribosomes translating the new
CC   HA sequence will usually terminate before the hydrophobic sequence
CC   is encountered, leading to the prodn. of a secreted form of the HA.
CC   It has been found that the primary sequence of the FRS can be rad-
CC   ically altered as long as the the secondary and tertiary structures
CC   are preserved, so it is possible to design an RFS which encodes
CC   hydrophobic amino acids, and therefore preserves the integrity of
CC   the anchor.
CC   See also AAR08418.
XX
SQ   Sequence   566 AA;

  Query Match             4.7%;  Score 110;  DB 11;  Length 566;
  Best Local Similarity   20.8%;  Pred. No. 1.7;
  Matches   97;  Conservative   67;  Mismatches  171;  Indels  132;  Gaps   24;

Qy      23 SAANAYMINSDMSDY------LSAVSDNFAERICSQVPKGS-----NCSASVSAYMSRCA 71
Db     101 NSENGICYPGDFIDYEELREQLSSVSS--FERF-EIFPKESSWPNHNTTKGVTAACSHAG 157

Qy      72 KQD------CLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDA-------NPANST 118
Db     158 KSSFYRNLLWLTEKEGSYP-------------KLKNSYVNKKGKEVLVLWGIHHPSNSK 203

Qy     119 EKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVD 178
Db     204 DQQ---NIYQNEN-AYVSVVTSNYNRRFTPEIAERPKVRDQAGRMNYYWTLLKPGDTIIF 259

Qy     179 EFGASFF-NKLSFTTGL-FGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSS--YKDY 234
Db     260 EANGNLIAPRYAFALSRGFGSG-------IITSNASMHECNTKCQTPLGAINSSLPFQNI 312

Qy     235 MDTQIPALPKFAKRFSLMVV-----------QRLLATVAGYVDTPW---YKKWY------ 274
Db     313 HPVTIGECPKYVRSAKLRMVTGLRNIPSIQSRGLFGAIAGFIEGGWTGMIDGWYGYHHQN 372
```

```
Qy    275 ------------MKLKNFMVNRV----------FIPTKKFFNKEIREPSKALKEKVSTD 311
                      | : |:|          |    | ||| : :  : | :||
Db    373 EQGSGYAADQKSTQNAINGITNKVNSVIEKMNIQFTAVGKEFNK-LEKRMENLNKKVDDG 431

Qy    312 TKD----------LFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTVD 361
          |          | ||   : |:|| :  :::  : :| :: |:||:     || |  :
Db    432 FLDIWTYNAELLVLLEN---ERTLDFHDSNVKNLYEKVKSQLKNNAKE-----IGNGCFE 483

Qy    362 F---INNEIRDPSKALIRKVSTGAEDL----FENKIGQGTVDFINNE 401
          |    :||        :  |  |  |      |:|: :  || :  |
Db    484 FYHKCDNE-------CMESVRNGTYDYPKYSEESKLNREKVDGVKLE 523
```

RESULT 20
ABG06855
ID ABG06855 standard; Protein; 2211 AA.
XX
AC ABG06855;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #6846.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS71042.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 37214; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2211 AA;

  Query Match 4.6%; Score 109.5; DB 22; Length 2211;
  Best Local Similarity 19.6%; Pred. No. 12;
  Matches 73; Conservative 61; Mismatches 130; Indels 109; Gaps 17;

```
Qy    174 YTNVDEFGASFFNKLSFTTG-----LFGWGIKRALKQIIRSNL-PLDIGTEHSVSRLQHI 227
          |||  :  :   ::| ||         |  : | :| : : |  ||    :   ::  :|
Db   1771 YTNNRQTESQIMSELPFTIASKTIKYLGIQLTRDVKDLFKENYKPLLNEIKEDTNKWNNI 1830

Qy    228 TSSYKDYMD-TQIPALPKFAKRFSLMVVQ----------------------RLLATV-- 261
            |:   ::  ::  |||   ||| : ::                        |:  ::
Db   1831 PCSWVGRINIMKMAILPKVIYRFSAIAIKLPMTFFTELEKTTLKFIWNQKRARIAKSILS 1890

Qy    262 -----------------AGYVDTPWYKKWYMK----------------LKNFMVNRVF 286
                           |    | ||  ||                    : |:::
Db   1891 QKNKAGGIMLPDFKLYYKATVTKTAWY--WYQNRDIDQWSRTEPSEVMPHIYNYLIFDKL 1948

Qy    287 IPTKKF-----FNKEIREPSKALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPS-KALK 340
            ||:     |||   | :|:  |:  |       ||    : : |  :|  : | |:
Db   1949 DKNKKWGKDSLFNKWCWENWLAICRKLKLDPFLTPYTKINSRWIKYLN--VRPKTRKTLE 2006

Qy    341 EKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLFENK---IGQGTVDF 397
          | : |  :|     || |  ||::    :  ||:  |      || : |     : |:
Db   2007 ENLGNTIQD-----IGMGK-DFMS----ETPKAMATKAKIDKWDLIKLKSFCTAKETIIR 2056

Qy    398 INNEIRDPS-----------KALIRKVYTEADDLFENKIGQGTVDFINKEIRDPSKALI 445
          :|   | |:           | || ::| |   ::: |    | : | |  :| ::  :
Db   2057 VN---RQPTEWEKIFATYSSDKGLISRIYNELQQIYKKK----TNNPIKKWAKDMNRHFL 2109

Qy    446 RKVSTEADNLLEK 458
          :: :  |   ::|
Db   2110 KEDTYAAKKHVKK 2122
```

RESULT 21
AAY20011
ID AAY20011 standard; Protein; 524 AA.
XX
AC AAY20011;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein, t301.aa.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR WPI; 1999-189980/16.
DR N-PSDB; AAX61708.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX

```
PS   Claim 12; Page 159-160; 275pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   524 AA;

  Query Match             4.6%;  Score 109;  DB 20;  Length 524;
  Best Local Similarity   20.0%;  Pred. No. 1.8;
  Matches   85;  Conservative   81;  Mismatches  152;  Indels  108;  Gaps   20;

Qy      36 DYLSAVSDNFAERICSQVPKGSNCSASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPL--- 92
           ||  : || |:     :|   |:  : :|: :   : :   : : :|:  |   : |
Db     146 DYQKSKSDPFSLESPLEVKYSSSIISYISSKLDNLSSKSNESFEKIKFYSEDLNEYLEQI 205

Qy      93 --TLPDPYQLEAAFILFK-------------------ESDANPANSTEK---RFWMRFR 127
             : :   :::: ::::                    || |:| |   |      |
Db     206 ETAISNTESIDSSILVYEQLRDTFSRFEKSIVDILKGFESIADPINDHNKYISEISSNFE 265

Qy     128 RGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEFGASFFNK 187
             :: |:|: :  ||   |     ::|||||  |:   :  :    : ||::   :| :
Db     266 --ESVSFFYSIDKNLEIFNKVATINSTDIENIKSKVFDLNIV----FENVNK---NFADL 316

Qy     188 LSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALP---- 243
           || |  |    : : |  |      | :      ::     |:   :  :|  |
Db     317 LSQTNSL--QSVNKLLVSISAQTNMLAMNAAIEAAKAGDAGKSFA-VVAEEIRKLAINSG 373

Qy     244 KFAK--RFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFM-----------VNRVFIPT 289
           |::|  :  |  |  ::| :   :|| :        |||:             :| :
Db     374 KYSKTIKDELKTVDSIIAVINSEIDTIY--------KNFIDIQDNVDNNFSRHEKVDLTL 425

Qy     290 KKFFNKEIREPSKALKEK-VSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKE------- 341
            | | ||| |     ||: :| |||                  |||     ||
Db     426 AKHF-KEIGE----FKERYLSHDTK-----------------IRDAKNMYKEIFNNHYF 462

Qy     342 ---KVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLFENKIGQGTVDFI 398
              | :| ::|| | |: :  :| ::: ::: |      | :  : : : |     :  |
Db     463 ISGKFNNFSQDLKEFKVSKMNLDAVSS-LQEYSSL----VKSSKDKILKTK---ELIQKI 514

Qy     399 NNEIRD 404
           |:||:|
Db     515 NDEIKD 520


RESULT 22
AAY20010
ID   AAY20010 standard; Protein; 553 AA.
XX
AC   AAY20010;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, f301.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   N-PSDB; AAX61707.
XX
PT   New isolated Borrelia burgdorferi nucleic acids  - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 12; Page 159; 275pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   553 AA;

  Query Match             4.6%;  Score 109;  DB 20;  Length 553;
  Best Local Similarity   20.0%;  Pred. No. 2;
  Matches   85;  Conservative   81;  Mismatches  152;  Indels  108;  Gaps   20;

Qy      36 DYLSAVSDNFAERICSQVPKGSNCSASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPL--- 92
           ||  : || |:     :|   |:  : :|: :   : :   : : :|:  |   : |
Db     175 DYQKSKSDPFSLESPLEVKYSSSIISYISSKLDNLSSKSNESFEKIKFYSEDLNEYLEQI 234

Qy      93 --TLPDPYQLEAAFILFK-------------------ESDANPANSTEK---RFWMRFR 127
             : :   :::: ::::                    || |:| |   |      |
Db     235 ETAISNTESIDSSILVYEQLRDTFSRFEKSIVDILKGFESIADPINDHNKYISEISSNFE 294

Qy     128 RGKNHSYFHDLVFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEFGASFFNK 187
             :: |:|: :  ||   |     ::|||||  |:   :  :    : ||::   :| :
Db     295 --ESVSFFYSIDKNLEIFNKVATINSTDIENIKSKVFDLNIV----FENVNK---NFADL 345

Qy     188 LSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALP---- 243
           || |  |    : : |  |      | :      ::      |:   :  :|  |
Db     346 LSQTNSL--QSVNKLLVSISAQTNMLAMNAAIEAAKAGDAGKSFA-VVAEEIRKLAINSG 402

Qy     244 KFAK--RFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFM-----------VNRVFIPT 289
           |::|  :  |  |  ::| :   :|| :        |||:             :| :
Db     403 KYSKTIKDELKTVDSIIAVINSEIDTIY--------KNFIDIQDNVDNNFSRHEKVDLTL 454

Qy     290 KKFFNKEIREPSKALKEK-VSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKE------- 341
            | | ||| |     ||: :| |||                  |||     ||
Db     455 AKHF-KEIGE----FKERYLSHDTK-----------------IRDAKNMYKEIFNNHYF 491

Qy     342 ---KVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLFENKIGQGTVDFI 398
              | :| ::|| | |: :  :| ::: ::: |      | :  : : : |     :  |
Db     492 ISGKFNNFSQDLKEFKVSKMNLDAVSS-LQEYSSL----VKSSKDKILKTK---ELIQKI 543

Qy     399 NNEIRD 404
           |:||:|
Db     544 NDEIKD 549


RESULT 23
ABG14594
ID   ABG14594 standard; Protein; 1054 AA.
XX
AC   ABG14594;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #14585.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
```

```
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS78781.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 44953; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1054 AA;

  Query Match             4.6%;  Score 109;  DB 22;  Length 1054;
  Best Local Similarity   20.2%;  Pred. No. 4.8;
  Matches  103;  Conservative   56;  Mismatches  158;  Indels  192;  Gaps   26;

Qy      50 CSQVPKGSNCSASVSAY----------MSRCAKQD----CLTLQSLKYP---------- 84
           || : :  : :| |:::          || |  ::    |     |: |
Db     439 CSNMKQCVDSNAYVASFPFGQCMEWYTMSTCPPENCSGYCTCSHCLEQPGCGWCTDPSNT 498

Qy      85 -----LEAKYQ-PLTLPD--------PYQLEAAFILFKESDAN----------PANSTE 119
                :|  |: |: :|         |  |  : :  ::|  |           | | |
Db     499 GKGKCIEGSYKGPVKMPSQAPTGNFYPQPLLNSSMCLEDSRYNWSFIHCPENKIPRNPTY 558

Qy     120 K----------RFWMRFRRG----KNHSYFHDLVFNLLEKNVTRDADATDIENFASRYL 164
           |              | :||    : ||   |   |:::|
Db     559 KGREGPLQGELQTTAQRNKRGHKQMEEHSMLMDRRINMVKK------------------ 599

Qy     165 YMATLYYKTYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRL 224
           : |:  |   :   :|| :|  ||  | |  |||
Db     600 ----VIYRFNTIPIKLPMTFFTELEKTTLKFIWNQKRA--------------------- 633

Qy     225 QHITSSYKDYMDTQIPA----LPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNF 280
           ||    | :  :  |    || |   :   | :           | ||  ||

Db     634 -HIA---KSILSQKNKAGGITLPDFRLYYKATVTK----------TAWY--WYQNRDLD 676

Qy     281 MVNRV----FIP-------------TKKF-----FNKEIREPSKALKEKVSTDTKDLFEN 318
             ||      ||               ||:     |||   |   |:  |:  |
Db     677 QRNRTEPSEIIPHIYNYLIFDKPDKNKKWGKDFLFNKWCWENWLAICRKLKLDPFLTLYT 736

Qy     319 KIGQGTVDFFNKEIRDPS-KALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRK 377
           ||    :   |  :|  : | |:| : |  :|      | |  ||:        ||:: |
Db     737 KINSRWIKDLN--VRPKTIKTLEENLGNTIQD-----TGMGK-DFMTK----TPKAMVTK 784

Qy     378 VSTGAEDLFENK---IGQGTVDFINNEIRDPS-----------KALIRKVYTEADDLFE 422
                || : |     : |   :|   | |:            | || ::| |   :::
Db     785 AKIDKWDLIKLKSFCTAKETTIRVN---RKPTEWEKIFATYSSDKGLISRIYKELKQIYK 841

Qy     423 NKIGQGTVDFINKEIRDPSKALIRKVSTE 451
             |       |  |:  :| : |  | |
Db     842 KK--------TNNPIKKWAKDMNRHFSKE 862


RESULT 24
ABG12107
ID   ABG12107 standard; Protein; 1104 AA.
XX
AC   ABG12107;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #12098.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS76294.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 42466; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
```

```
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1104 AA;

  Query Match             4.6%;  Score 109;  DB 22;  Length 1104;
  Best Local Similarity   20.2%;  Pred. No. 5.1;
  Matches  103;  Conservative   56;  Mismatches  158;  Indels  192;  Gaps   26;

Qy      50 CSQVPKGSNCSASVSAY----------MSRCAKQD----CLTLQSLKYP---------- 84
           || : :  : :| |:::           || |  ::     |      |: |
Db     439 CSNMKQCVDSNAYVASFPFGQCMEWYTMSTCPPENCSGYCTCSHCLEQPGCGWCTDPSNT 498

Qy      85 -----LEAKYQ-PLTLPD--------PYQLEAAFILFKESDAN----------PANSTE 119
                :|  |: |: :|          |  |  : :  ::|  |           | | |
Db     499 GKGKCIEGSYKGPVKMPSQAPTGNFYPQPLLNSSMCLEDSRYNWSFIHCPENKIPRNPTY 558

Qy     120 K----------RFWMRFRRG----KNHSYFHDLVFNLLEKNVTRDADATDIENFASRYL 164
           |          | :||    : ||   |   |:::|
Db     559 KGREGPLQGELQTTAQRNKRGHKQMEEHSMLMDRRINMVKK------------------ 599

Qy     165 YMATLYYKTYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDIGTEHSVSRL 224
           : |:  |   :   :|| :|  ||  | |  |||
Db     600 ----VIYRFNTIPIKLPMTFFTELEKTTLKFIWNQKRA---------------------- 633

Qy     225 QHITSSYKDYMDTQIPA----LPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNF 280
            ||    |  :  :  |    || |   :   | :           | ||  ||
Db     634 -HIA---KSILSQKNKAGGITLPDFRLYYKATVTK----------TAWY--WYQNRDLD 676

Qy     281 MVNRV----FIP-------------TKKF-----FNKEIREPSKALKEKVSTDTKDLFEN 318
             ||      ||              ||:     |||   |   |:  |:  |
Db     677 QRNRTEPSEIIPHIYNYLIFDKPDKNKKWGKDFLFNKWCWENWLAICRKLKLDPFLTLYT 736

Qy     319 KIGQGTVDFFNKEIRDPS-KALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRK 377
           ||    :   |  :|  : | |:| : |  :|      | |. ||:         ||:: |
Db     737 KINSRWIKDLN--VRPKTIKTLEENLGNTIQD-----TGMGK-DFMTK----TPKAMVTK 784

Qy     378 VSTGAEDLFENK---IGQGTVDFINNEIRDPS------------KALIRKVYTEADDLFE 422
                || : |     : |   :|   | |:            | || ::| |   :::
Db     785 AKIDKWDLIKLKSFCTAKETTIRVN---RKPTEWEKIFATYSSDKGLISRIYKELKQIYK 841

Qy     423 NKIGQGTVDFINKEIRDPSKALIRKVSTE 451
           |         |  |:  :| : |  | |
Db     842 KK--------TNNPIKKWAKDMNRHFSKE 862


RESULT 25
ABB61247
ID   ABB61247 standard; Protein; 2470 AA.
XX
AC   ABB61247;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 10533.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL05350.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 10533; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2470 AA;

  Query Match             4.6%;  Score 108;  DB 22;  Length 2470;
  Best Local Similarity   21.9%;  Pred. No. 18;
  Matches   82;  Conservative   45;  Mismatches  157;  Indels   90;  Gaps   12;

Qy     100 LEAAFILFKESDANPANSTEKRFWMRFRRGK-NHSYFHDLVFNLLEKNVTRD-------- 150
           | || |:  : ::   :|  :  | |    : | |:  ||  :  :| ||||
Db     200 LRAALIVTAQRESTKQSSEPQ--WYRICYDEANGSFNADLGSSKDQKGVTRDDRIHGGLV 257

Qy     151 --------ADATDIENFASRYLYMATLYYKTYTN-----------------VDEFGASF 184
                   |:||    :  ||  : ||: ||  |                 |      |
Db     258 VFNELFRCANAT----WERRYTSLKTLFPKTQHNKFLEASSSSSMGSQLNTLVPRLKVPF 313

Qy     185 FNKLSFTTGLFGWG------IKRALKQIIRSNLPLDIGTEHSVS----RLQHITSSYKDY 234
           :|| |    | |       | |   :: |    :|  ||  |     |:  ||
Db     314 IDKLGSTQTHLGEGEHHKGVAKFASHNVLESAYAQEILQEHYTSICDNVLEQRTSKSPYV 373

Qy     235 MDTQIPALPKFAKRFSLMVVQRLLAT--------------------VAGYVDTPWYKKW 273
                :  ||: |     : |:: | |                        ||:
Db     374 QQALLQILPRLAAFNRAVFVEKYLQTCVSHLMQILRGKEKDRTVAYITIGYMAVAVQSAI 433

Qy     274 YMKLKNFMVN-RVFIPTKKFFNKE----------IREPSKALKEKVSTDTKDLFENKIG 321
           : | : | : :| :|:|   :|           |   : |:| ::: | ||: |
Db     434 EVHLSSIMTSVKVALPSKDLTSKRKVPVDPAVFACITLLAHAVKSEIADDVKDILEQMFY 493

Qy     322 QGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTG 381
           |        :|:    | | |      : |  ||  :   :|     |  ||      |
Db     494 TGLSPALTVCLRE----LSENVPQLKSAITEGLIGILSQVLMNKAAILPYTALPTIAIDG 549

Qy     382 AEDLFENKIGQGTV 395
           :  | :|  |  ||
Db     550 S--LMQNGDGATTV 561


RESULT 26
AAG53797
ID   AAG53797 standard; Protein; 446 AA.
XX
AC   AAG53797;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 68525.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
```

XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.6%; Score 107.5; DB 21; Length 446;
Best Local Similarity 18.3%; Pred. No. 2;
Matches 77; Conservative 74; Mismatches 112; Indels 157; Gaps 22;

```
Qy    137 DLVFNLLEKNVTRD--ADATDI---ENFASRYLYMATL---------------------- 169
          |  | :| |    |   :| ||     :  || ::  :
Db     40 DRTFGVLTKIDLMDQGTNAVDILEGRGYKLRYPWVGVVNRSQADINKSVDMIAARRRERD 99

Qy    170 YYKT---YTNVDE-FGASFFNKLSFTTGLFGWGIKRALKQIIRSNLP----------LDI 215
          |::|   | :: |  |: :  |:          : : |: :|:| :|           ::
Db    100 YFQTSPEYRHLTERMGSEYLGKM----------LSKHLEVVIKSRIPGLQSLITKTISEL 149

Qy    216 GTEHSVSRL------------------QHITSSYKDYM--------------DTQIPAL 242
           ||  :|||                 :  ::|:::               | | ||
Db    150 ETE--LSRLGKPVAADAGGKLYMIMEICRAFDQTFKEHLDGTRSGGEKINSVFDNQFPAA 207

Qy    243 PK---FAKRFSLMVVQRLLATVAGY---VDTPWYKKWYMKL-KNFMVNRVFIPTKKFFNK 295
          |   | |  |:  |::|:    ||   :  |  :: | :| :: :|:
Db    208 IKRLQFDKHLSMDNVRKLITEADGYQPHLIAP--EQGYRRLIESCLVS------------ 253

Qy    296 EIREPSKALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKD------ 349
           || |::|  : | :  |||    :|:        |::     |: :||  | |
Db    254 -IRGPAEAAVDAVHSILKDLIHKSMGE------TSELKQ-YPTLRVEVSGAAVDSLDRMR 305

Qy    350 ---------LFENKIGQGTVDFINNEIRDPSKA--LIRKVSTGAEDLFENKIGQGTVDFI 398
                  | : : |  ||:|     :|  |       :     | :  :||   : ::
Db    306 DESRKATLLLVDMESGYLTVEFFRKLPQDSEKGGNPTHSIFDRYNDAYLRRIGSNVLSYV 365

Qy    399 NN----------------EIRDPSKALIRKVYTEADDLFENKIGQGTVDFINKEI-RDPS 441
          |                 ::|:  ::|:        |:|  ::||  :  ::| :  ||:
Db    366 NMVCAGLRNSIPKSIVYCQVREAKRSLL--------DIFFTELGQKEMSKLSKLLDEDPA 417
```

RESULT 27
AAG53796
ID AAG53796 standard; Protein; 522 AA.
XX
AC AAG53796;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 68524.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.6%; Score 107.5; DB 21; Length 522;
Best Local Similarity 18.3%; Pred. No. 2.4;
Matches 77; Conservative 74; Mismatches 112; Indels 157; Gaps 22;

```
Qy    137 DLVFNLLEKNVTRD--ADATDI---ENFASRYLYMATL---------------------- 169
          |  | :| |    |   :| ||     :  || ::  :
Db    116 DRTFGVLTKIDLMDQGTNAVDILEGRGYKLRYPWVGVVNRSQADINKSVDMIAARRRERD 175

Qy    170 YYKT---YTNVDE-FGASFFNKLSFTTGLFGWGIKRALKQIIRSNLP----------LDI 215
          |::|   | :: |  |: :  |:          : : |: :|:| :|          ::
Db    176 YFQTSPEYRHLTERMGSEYLGKM----------LSKHLEVVIKSRIPGLQSLITKTISEL 225
```

```
Qy    216 GTEHSVSRL------------------QHITSSYKDYM-------------DTQIPAL 242
Db    226 ETE--LSRLGKPVAADAGGKLYMIMEICRAFDQTFKEHLDGTRSGGEKINSVFDNQFPAA 283

Qy    243 PK---FAKRFSLMVVQRLLATVAGY---VDTPWYKKWYMKL-KNFMVNRVFIPTKKFFNK 295
Db    284 IKRLQFDKHLSMDNVRKLITEADGYQPHLIAP--EQGYRRLIESCLVS------------ 329

Qy    296 EIREPSKALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKD------ 349
Db    330 -IRGPAEAAVDAVHSILKDLIHKSMGE------TSELKQ-YPTLRVEVSGAAVDSLDRMR 381

Qy    350 ---------LFENKIGQGTVDFINNEIRDPSKA--LIRKVSTGAEDLFENKIGQGTVDFI 398
Db    382 DESRKATLLLVDMESGYLTVEFFRKLPQDSEKGGNPTHSIFDRYNDAYLRRIGSNVLSYV 441

Qy    399 NN---------------EIRDPSKALIRKVYTEADDLFENKIGQGTVDFINKEI-RDPS 441
Db    442 NMVCAGLRNSIPKSIVYCQVREAKRSLL--------DIFFTELGQKEMSKLSKLLDEDPA 493
```

RESULT 28
AAG53795
ID AAG53795 standard; Protein; 610 AA.
XX
AC AAG53795;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 68523.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.6%; Score 107.5; DB 21; Length 610;
Best Local Similarity 18.3%; Pred. No. 3;
Matches 77; Conservative 74; Mismatches 112; Indels 157; Gaps 22;

```
Qy    137 DLVFNLLEKNVTRD--ADATDI---ENFASRYLYMATL---------------------- 169
          |  | :| |    |   :| ||     :  || ::  :
Db    204 DRTFGVLTKIDLMDQGTNAVDILEGRGYKLRYPWVGVVNRSQADINKSVDMIAARRRERD 263

Qy    170 YYKT---YTNVDE-FGASFFNKLSFTTGLFGWGIKRALKQIIRSNLP----------LDI 215
          |::|   | :: |  |: :  |:          : : |: :|:| :|          ::
Db    264 YFQTSPEYRHLTERMGSEYLGKM----------LSKHLEVVIKSRIPGLQSLITKTISEL 313

Qy    216 GTEHSVSRL------------------QHITSSYKDYM--------------DTQIPAL 242
           ||  :|||                   :    ::|:::              | | ||
Db    314 ETE--LSRLGKPVAADAGGKLYIMEICRAFDQTFKEHLDGTRSGGEKINSVFDNQFPAA 371

Qy    243 PK---FAKRFSLMVVQRLLATVAGY---VDTPWYKKWYMKL-KNFMVNRVFIPTKKFFNK 295
          |   | |  |:  |::|:    ||   :  |  :: | :| :: :|:
Db    372 IKRLQFDKHLSMDNVRKLITEADGYQPHLIAP--EQGYRRLIESCLVS------------ 417

Qy    296 EIREPSKALKEKVSTDTKDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKD------ 349
           || |::|  : | :  |||    :|:        |::     |: :||  | |
Db    418 -IRGPAEAAVDAVHSILKDLIHKSMGE------TSELKQ-YPTLRVEVSGAAVDSLDRMR 469

Qy    350 ---------LFENKIGQGTVDFINNEIRDPSKA--LIRKVSTGAEDLFENKIGQGTVDFI 398
                  | : : |  ||:|     :|  |       :     | :  :||   : ::
Db    470 DESRKATLLLVDMESGYLTVEFFRKLPQDSEKGGNPTHSIFDRYNDAYLRRIGSNVLSYV 529

Qy    399 NN---------------EIRDPSKALIRKVYTEADDLFENKIGQGTVDFINKEI-RDPS 441
          |               ::|:  ::|:        |:|  ::||  :  ::| :  ||:
Db    530 NMVCAGLRNSIPKSIVYCQVREAKRSLL--------DIFFTELGQKEMSKLSKLLDEDPA 581
```

RESULT 29
AAG79240
ID AAG79240 standard; Protein; 714 AA.
XX
AC AAG79240;
XX
DT 03-JAN-2002 (first entry)
XX
DE Amino acid sequence of S-layer protein of C. difficile strain 17.
XX
KW Surface layer protein; S-layer protein; pseudomembranous colitis; PMC;
KW cell wall protein; gastrointestinal illness; abscess; wound infection;
KW osteomyelitis; urogenital tract infection; septicemia; peritonitis;
KW pleuritis.
XX
OS Clostridium difficile.
XX
PN WO200173040-A1.
XX
PD 04-OCT-2001.
XX
PF 23-MAR-2001; 2001WO-GB01305.
XX
PR 24-MAR-2000; 2000GB-0007263.
XX
PA (UNLO) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX
PI Fairweather NF, Calabi E;
XX
DR WPI; 2001-616508/71.
DR N-PSDB; AAI65839.
XX
PT Novel polypeptides and polynucleotides of cell wall proteins of
PT Clostridium difficile especially S-layer cell wall protein useful for
PT preventing and treating the infection caused by the bacteria -

```
XX
PS   Claim 1; Page 56; 62pp; English.
XX
CC   The present sequence represents a surface layer (S-layer) protein of
CC   Clostridium difficile. The S-layer proteins are the predominant cell
CC   wall protein. There are two distinct S-layer proteins in C. difficile,
CC   a 45 kDa and 36 kDa protein. S-layer polypeptides and polynucleotides
CC   are useful for treating and/or preventing a disease associated with
CC   C.difficile infection in a subject. Such diseases include
CC   pseudomembranous colitis (PMC) in humans characterized by diarrhoea, a
CC   severe inflammation of the colonic mucosa, and formation of
CC   pseudomembranes that are composed of fibrin, mucus, necrotic epithelial
CC   cells and leukocytes; gastrointestinal illness, abscesses, wound
CC   infections, osteomyelitis, urogenital tract infections, septicemia,
CC   peritonitis, and pleuritis.
XX
SQ   Sequence   714 AA;

  Query Match             4.5%;  Score 106.5;  DB 22;  Length 714;
  Best Local Similarity   21.5%;  Pred. No. 4.5;
  Matches  102;  Conservative   81;  Mismatches  177;  Indels  115;  Gaps   26;

Qy      15 LAASESVD---SAANAYMINSDMSDYLSAVSDNFAERICSQVPKGSNCSASVSAYMSRCA 71
           :||: ::    ::|    |: | | |:||  :| |::      : |    :: | :
Db     257 MAATGTIKVRVTSAKEESIDVDSSSYISA--ENLAKKYVFNPKEVSEAYNAIVALQNDGI 314

Qy      72 KQDCLTLQSLKYPLEAKYQPLTLPDPYQLE--AAFILFKESDA-NPANSTEKRFWMRFRR 128
           : | : |      :  ||| :  |:  :||  :| |:   :|| :||  | |   ::  :
Db     315 ESDLVQL------VNGKYQVIFYPEGKRLETKSADII---ADADSPAKITIKANKLKDLK 365

Qy     129 GKNHSYFHDL-VFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTNVDEFGA----- 182
                |  ||  :|    ||   |    ||      : ::: ||    | |:  |
Db     366 ----DYVDDLKTYNNTYSNVVTVAGEDRIET----AIELSSKYY----NSDDKNAITDDA 413

Qy     183 ----SFFNKLSFTTGLFGWGI--KRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMD 236
                   |   ||    :  ::    :: |   ||   :  : |: :: |      |
Db     414 VNNIVLVGSTSIVDGLVASPLASEKTAPLLLTSKDKLDSSVKSEIKRVMNLKS------D 467

Qy     237 TQIPALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKE 296
           | |    |               :|| |::   |    :|||     : :   :   ::
Db     468 TGINTSKK-------------VYLAGGVNS-ISKDVENELKN-----MGLKVTRLSGED 507

Qy     297 IREPSKALKEKVSTDTKDLFENKIGQGTVDFFN-----KEIRD---------PSKALKEK 342
             | | |: :::  |    |    | |  |  :      :::|           ||  ::
Db     508 RYETSLAIADEIGLDNDKAFVVG-GTGLADAMSIAPVASQLKDGDATPIVVVDGKA--KE 564

Qy     343 VSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLFENKIGQGTVDFINNEI 402
           :|:|||      :|   || |  :           ||   |:  ::  |: | | |: :
Db     565 ISDDAKSF----LGTSDVDIIGGK---------NSVSKEIEESIDSATGK-TPDRISGDD 610

Qy     403 RDPSKALIRKVYTEADDLFENKIGQGTVDFINKEIRDPSKALIRKVSTEADNLLE 457
           |  : | : |     || |  | |:    |: |:           ||: | |::
Db     611 RQATNAEVLK----EDDYF--KDGEVVNYFVAKD----------GSTKEDQLVD 648


RESULT 30
AAB50676
ID   AAB50676 standard; Protein; 1508 AA.
XX
AC   AAB50676;
XX
DT   19-MAR-2001  (first entry)
XX
DE   C. elegans alpha-2-macroglobulin cDNA sequence SEQ ID NO:70.
XX
KW   Human; Caenorhabditis elegans; UNC-5; splice variant; nematode worm;
KW   protein-protein interaction; identification.
XX
OS   Caenorhabditis elegans.
XX
PN   WO200073328-A2.
XX
PD   07-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-EP05108.
XX
PR   01-JUN-1999;   99GB-0012755.
XX
PA   (DEVG-) DEVGEN NV.
XX
PI   Van Criekinge W,  Roelens I,  Bogaert T,  Verwaerde P;
XX
DR   WPI; 2001-016508/02.
XX
PT   Three variants of human unc-5C cDNAs (unc-5Cb, unc-5Cc and unc-5C8) and
PT   a human unc-5HS1 cDNA, useful in yeast two hybrid experiments for
PT   identifying unknown human cDNAs which encode proteins that interact
PT   with the human unc-5C protein -
XX
PS   Example 4; Page 197-202; 246pp; English.
XX
CC   The present invention describes 3 variants of human unc-5C cDNAs
CC   (unc-5Cb, unc-5Cc and unc-5C8) which correspond to alternatively spliced
CC   unc-5C transcripts, and a human unc-5HS1 cDNA which shares homology with
CC   the Rattus norvegicus unc-5HS1 cDNA. Also described are assays based on
CC   protein-protein-interactions between the unc-5 protein and a variety of
CC   different interacting proteins. The unc-5C variant cDNAs and unc-5HS1
CC   cDNA are useful in methods for identifying compounds which reduce or
CC   inhibit the lethal phenotype associated with the expression of the
CC   unc-5 death domain in yeast. They are also useful in yeast two hybrid
CC   experiments for identifying unknown human cDNAs which encode proteins
CC   that interact with the human unc-5C protein. AAC90914 to AAC90971 and
CC   AAB50646 to AAB50693 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence   1508 AA;

  Query Match             4.5%;  Score 106;  DB 22;  Length 1508;
  Best Local Similarity   21.7%;  Pred. No. 14;
  Matches   62;  Conservative   53;  Mismatches  117;  Indels   54;  Gaps   13;

Qy     105 ILFKESDANPANST-EKRFWMRFRRGKNHSYFHDLVFNLLEKNVTR----DADATDI--- 156
           :  |: |  |  :|  ||  ::     |: | ||   :| |:  |:    ||  | :
Db     394 VALKQMDDTPVKATLPKR--VQVSTFYNYPYNHD-TSSLQEEKETKIVEVDAHGTSVLTL 450

Qy     157 --------------------ENFASRYLYMATLYYKTYTNVDEFGASFFNKLSFTTGLFG 196
                               :|| :  :| ::|| :   :  :   ||   |:   |
Db     451 QPPINCTSARIEAHYDIGGKDNFTATPIY-SSLYVEAAVSPTK---SFLQLLADNEGAVD 506

Qy     197 WGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSY----KDYMDTQIPALPKFAKRFSLM 252
                |  ::|   :::  ||   |   :||   : |        ::     ||    | :  |:
Db     507 VG--KSLSFSLKATQPLSTITYQVMSRSNIVVSQQMTVNSEHATISFPATANMAPKSRLI 564

Qy     253 VVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREPSKALKEKVSTDT 312
           |   : ::    ||   :     |::    |:| :      :|:  || : :| ||::|
Db     565 VYAIIESSQEVLVDALDF-----KVEGIFQNQVALS----IDKQAVEPGQNVKFKVTSD- 614

Qy     313 KDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQG 358
           |: |   :|   ||      ::  :   :|||  | ::   | :| |
Db     615 KNSF---VGLLVVDQSVLLLKTGNDITREKVEQDLENYDSNNVGGG 657


RESULT 31
AAB50677
ID   AAB50677 standard; Protein; 1519 AA.
XX
AC   AAB50677;
XX
DT   19-MAR-2001  (first entry)
XX
DE   C. elegans alpha-2-macroglobulin protein sequence SEQ ID NO:71.
XX
KW   Human; Caenorhabditis elegans; UNC-5; splice variant; nematode worm;
KW   protein-protein interaction; identification.
XX
OS   Caenorhabditis elegans.
```

```
XX
PN   WO200073328-A2.
XX
PD   07-DEC-2000.
XX
PF   02-JUN-2000; 2000WO-EP05108.
XX
PR   01-JUN-1999;   99GB-0012755.
XX
PA   (DEVG-) DEVGEN NV.
XX
PI   Van Criekinge W,  Roelens I,  Bogaert T,  Verwaerde P;
XX
DR   WPI; 2001-016508/02.
XX
PT   Three variants of human unc-5C cDNAs (unc-5Cb, unc-5Cc and unc-5C8) and
PT   a human unc-5HS1 cDNA, useful in yeast two hybrid experiments for
PT   identifying unknown human cDNAs which encode proteins that interact
PT   with the human unc-5C protein -
XX
PS   Example 4; Page 202-207; 246pp; English.
XX
CC   The present invention describes 3 variants of human unc-5C cDNAs
CC   (unc-5Cb, unc-5Cc and unc-5C8) which correspond to alternatively spliced
CC   unc-5C transcripts, and a human unc-5HS1 cDNA which shares homology with
CC   the Rattus norvegicus unc-5HS1 cDNA. Also described are assays based on
CC   protein-protein-interactions between the unc-5 protein and a variety of
CC   different interacting proteins. The unc-5C variant cDNAs and unc-5HS1
CC   cDNA are useful in methods for identifying compounds which reduce or
CC   inhibit the lethal phenotype associated with the expression of the
CC   unc-5 death domain in yeast. They are also useful in yeast two hybrid
CC   experiments for identifying unknown human cDNAs which encode proteins
CC   that interact with the human unc-5C protein. AAC90914 to AAC90971 and
CC   AAB50646 to AAB50693 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence   1519 AA;

  Query Match             4.5%;  Score 106;  DB 22;  Length 1519;
  Best Local Similarity   21.7%;  Pred. No. 14;
  Matches   62;  Conservative   53;  Mismatches  117;  Indels   54;  Gaps   13;

Qy     105 ILFKESDANPANST-EKRFWMRFRRGKNHSYFHDLVFNLLEKNVTR----DADATDI--- 156
           :  |: |  |  :|  ||  ::     |: | ||   :| |:  |:    ||  | :
Db     394 VALKQMDDTPVKATLPKR--VQVSTFYNYPYNHD-TSSLQEEKETKIVEVDAHGTSVLTL 450

Qy     157 --------------------ENFASRYLYMATLYYKTYTNVDEFGASFFNKLSFTTGLFG 196
                               :|| :  :| ::|| :   :  :   ||   |:   |
Db     451 QPPINCTSARIEAHYDIGGKDNFTATPIY-SSLYVEAAVSPTK---SFLQLLADNEGAVD 506

Qy     197 WGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSY----KDYMDTQIPALPKFAKRFSLM 252
           |  ::|   :::  ||   |   :||   : |       ::     ||    | :  |:
Db     507 VG--KSLSFSLKATQPLSTITYQVMSRSNIVVSQQMTVNSEHATISFPATANMAPKSRLI 564

Qy     253 VVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRVFIPTKKFFNKEIREPSKALKEKVSTDT 312
           |   : ::    ||   :     |::    |:| :      :|:  || : :| ||::|
Db     565 VYAIIESSQEVLVDALDF-----KVEGIFQNQVALS----IDKQAVEPGQNVKFKVTSD- 614

Qy     313 KDLFENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQG 358
           |: |   :|   ||     ::  :   :|||  | ::   | :| |
Db     615 KNSF---VGLLVVDQSVLLLKTGNDITREKVEQDLENYDSNNVGGG 657


RESULT 32
ABG08970
ID   ABG08970 standard; Protein; 541 AA.
XX
AC   ABG08970;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #8961.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS73157.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 39329; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   541 AA;

  Query Match             4.5%;  Score 105.5;  DB 22;  Length 541;
  Best Local Similarity   18.7%;  Pred. No. 3.7;
  Matches  103;  Conservative   75;  Mismatches  185;  Indels  189;  Gaps   25;

Qy      30 INSDMSDYLSAVSD-NFAERICSQVPKGSN---CSASVSAY-----------MSRCAKQ 73
           :| |: :  ||:   :  :   :  || :     |||   |            :|:| :
Db      19 VNKDIQELNSALHQADLIDIYRTLHPKSTEYTFFSASHHTYSKIDRIVGSKTLLSKCKRT 78

Qy      74 DCLT---LQSLKYPLEAKYQPLT-------------LPDPYQLEAAFILFKESDANPANS 117
           : ||   |      || : | ||              | | ::|   |    ::|   |
Db      79 EVLTNCLLDHSAIKLELRIQKLTQNRSTTWKLNNLLLNDYWKL--VFEKINKTDRPLARP 136

Qy     118 TEKRFWMRFRRGK----------------------NHSYFHDLVFNLLE---------- 144
            :|     :| |                      |  |:  |  | ||
Db     137 IKK------KREKNQIDAIKNDKGDITTNPTEIQTTNREYYKHLYANKLENLEEMHKFLD 190

Qy     145 ------------KNVTRDADATDIENFASRYLYMATLYYKTYTNVDEFGASFFNKLSFTT 192
                       ::: |    ::||        : :|  |     | | | |: :: :
Db     191 TYTSPRLNQEEVESLNRSITGSEIE------AIIKSLTTKKSPEPDGFTAEFYQRIKY-- 242

Qy     193 GLFGWGIKRALKQIIRSNL-PLDIGTEHSVSRLQHITSSYKDYMD-TQIPALPKFAKRFS 250
```

```
Db    243 --LGIQLTRDVKDLFKENYKPLLNEIKEDTNKWKNIPCSWIGRINIVKMAILPKVIYRFN 300

Qy    251 LMVVQ-------RLLATVAGYV----------------------------------DT 367

Db    301 AIPIKLPMTFFTELEKTTLKFIWNQKRACIAKTMLSIKNKVRGITLPDFKLYYKATVTKT 360

Qy    268 PWYKKWYMK--------------LKNFMVNRVFIPTKK--------FFNKEIREPSKALK 305

Db    361 AWY--WYQNRDIDQGNRTEPSEIIPHFYNHLIFDKPDKNKKWGKDSLFNKCCWENWLAIC 418

Qy    306 EKVSTDTKDLFENKIGQGTVDFFNKEIRDPS-KALKEKVSNDAKDLFENKIGQGTVDFIN 364

Db    419 RKLKLDPFLTPYTKINSRWIKDLN--VRPKTIKTLEENLGNTIQD-----IGMGK-DFMS 470

Qy    365 NEIRDPSKALIRKVSTGAEDLFENK---IGQGTVDFINNE---------IRDPSKALIRK 412

Db    471 K----TPKAMATKAKIDKSDLIKLKSFCTAKETTIRVNRQPTEWEKIFAIYSSDKGLICR 526

Qy    413 VYTEADDLFENK 424

Db    527 IYNELKQIYKKK 538
```

RESULT 33
AAW01670
ID AAW01670 standard; Protein; 572 AA.
XX
AC AAW01670;
XX
DT 19-AUG-1997 (first entry)
XX
DE Influenza A/Texas/36/91 recombinant haemagglutinin protein.
XX
KW primer; PCR; polymerase chain reaction; universal; amplify; HA;
KW haemagglutinin; recombinant production; baculovirus expression system;
KW vaccine; insect cell culture.
XX
OS Synthetic.
XX

FH	Key	Location/Qualifiers
FT	Peptide	1..18
FT		/label= AcNPV_61K_protein_signal_sequence
FT	Protein	19..554
FT		/label= mature_recombinant_haemagglutinin

XX
PN WO9637624-A1.
XX
PD 28-NOV-1996.
XX
PF 26-MAY-1995; 95WO-US06750.
XX
PR 26-MAY-1995; 95WO-US06750.
XX
PA (MICR-) MICROGENESYS INC.
PA (MGPM-) MG*PMC LLC.
XX
PI Hackett CS, Smith GE, Volvovitz F, Voznesensky AI;
PI Wilkinson BE;
XX
DR WPI; 1997-021228/02.
DR N-PSDB; AAT59213.
XX
PT Recombinant influenza haemagglutinin produced in baculovirus system
PT - avoids problems of growing virus in eggs and produces stable,
PT un-cleaved protein useful in vaccines
XX
PS Example 3; Page 73-74; 107pp; English.
XX
CC Recombinant influenza haemagglutinin (HA) expressed in a
CC baculovirus expression system in cultured insect cells, allows vaccine
CC production without the need to grow virus in eggs. A purer, less
CC allergenic product is obtained and antigen drift caused by passages
CC through eggs is avoided. There is no need for viral inactivation or
CC organic solvent extn. of viral membrane components and vaccines can be
CC prepd. rapidly and cost effectively from primary sources of infection.
CC Recombinant HA is more stable (esp. for B strains) than HA1/HA2 complexes
CC and maintain correct folding during purification and storage. The present
CC sequence shows the N-terminal end of the HA protein for influenza
CC A/Texas/36/91 (sequence range 1-481).
XX
SQ Sequence 572 AA;

Query Match 4.5%; Score 105.5; DB 18; Length 572;
Best Local Similarity 22.3%; Pred. No. 4;
Matches 60; Conservative 37; Mismatches 87; Indels 85; Gaps 14;

```
Qy    195 FGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSS--YKDYMDTQIPALPKFAKRFSLM 252

Db    286 FGSG-------IITSNASMDECDAKCQTPQGAINSSLPFQNVHPVTIGECPKYVRSTKLR 338

Qy    253 VV-----------QRLLATVAGYVDTPW---YKKWY------------------MKLKN 279

Db    339 MVTGLRNIPSIQSRGLFGAIAGFIEGGWTGMIDGWYGYHHQNEQGSGYAADQKSTQNAIN 398

Qy    280 FMVNRV----------FIPTKKFFNKEIREPSKALKEKVSTDTKD----------LFENK 319

Db    399 GITNKVNSVIEKMNTQFTAVGKEFNK-LERRMENLNKKVDDGFLDIWTYNAELLVLLEN- 456

Qy    320 IGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTVDF---INNEIRDPSKALIR 376

Db    457 -GR-TLDFHDSNVKNLYEKVKSQLKNNAKE-----IGNGCFEFYHKCNNE-------CME 502

Qy    377 KVSTGAEDL----FENKIGQGTVDFINNE 401

Db    503 SVKNGTYDYPKYSEESKLNRGKIDGVKLE 531
```

RESULT 34
AAW75442
ID AAW75442 standard; Protein; 572 AA.
XX
AC AAW75442;
XX
DT 13-APR-1999 (first entry)
XX
DE Influenza virus A/Texas/36/91 recombinant HA protein.
XX
KW Recombinant; glycosylation; influenza virus; haemagglutinin; baculovirus;
KW fusion protein; expression system; insect cell; immunogen; vaccine;
KW immune response; primer; PCR; amplification; reverse transcription;
KW human; bird.
XX
OS Synthetic.
OS Influenza virus.
XX

FH	Key	Location/Qualifiers
FT	Peptide	1..18
FT		/note= "AcNPV 61K signal peptide"
FT	Protein	19..554
FT		/note= "mature haemagglutinin protein"

XX
PN US5858368-A.
XX
PD 12-JAN-1999.
XX
PF 30-MAY-1995; 95US-0453848.
XX
PR 30-MAY-1995; 95US-0453848.
PR 13-SEP-1993; 93US-0120607.
XX
PA (PROT-) PROTEIN SCI CORP.
XX
PI Hackett CS, Smith GE, Volvovitz F, Voznesensky AI;
PI Wilkinson BE;
XX

```
DR   WPI; 1999-119782/10.
DR   N-PSDB; AAX00774.
XX
PT   Recombinant influenza virus haemagglutinin fusion protein - for use
PT   in vaccines against influenza
XX
PS   Example 3; Column 43-48; 50pp; English.
XX
CC   The invention relates to the production of a recombinant glycosylated
CC   influenza virus haemagglutinin fusion protein by a baculovirus expression
CC   system in cultured insect cells, where the protein is at least 95% pure,
CC   is immunogenic, induces a protective immune response when used as a
CC   vaccine, and comprises a second protein fused to the haemagglutinin.
CC   This sequence represents the recombinant haemagglutinin from the
CC   influenza virus type A strain Texas/36/91 linked to the baculovirus
CC   Autographa californica nuclear polyhedrosis virus (AcNPV) 61K protein
CC   signal sequence.  The vaccine is used for vaccinating animals (including
CC   humans) or birds against influenza.
XX
SQ   Sequence   572 AA;

  Query Match             4.5%;  Score 105.5;  DB 20;  Length 572;
  Best Local Similarity   22.3%;  Pred. No. 4;
  Matches   60;  Conservative   37;  Mismatches   87;  Indels   85;  Gaps   14;

Qy     195 FGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSS--YKDYMDTQIPALPKFAKRFSLM 252
           || |       || ||  :|       :    | ||  :::     |   ||: :   |
Db     286 FGSG-------IITSNASMDECDAKCQTPQGAINSSLPFQNVHPVTIGECPKYVRSTKLR 338

Qy     253 VV----------QRLLATVAGYVDTPW---YKKWY------------------MKLKN 279
           :|          : |   :||:::  |      ||                      |
Db     339 MVTGLRNIPSIQSRGLFGAIAGFIEGGWTGMIDGWYGYHHQNEQGSGYAADQKSTQNAIN 398

Qy     280 FMVNRV----------FIPTKKFFNKEIREPSKALKEKVSTDTKD----------LFENK 319
           : |:|          |    | ||| :    : | :||     |          | ||
Db     399 GITNKVNSVIEKMNTQFTAVGKEFNK-LERRMENLNKKVDDGFLDIWTYNAELLVLLEN- 456

Qy     320 IGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTVDF---INNEIRDPSKALIR 376
           |: |:|| :  :::  : :| :: |:||:     || |  :|    |||       :
Db     457 -GR-TLDFHDSNVKNLYEKVKSQLKNNAKE-----IGNGCFEFYHKCNNE-------CME 502

Qy     377 KVSTGAEDL----FENKIGQGTVDFINNE 401
           |  |  |      |:|: :| :| :  |
Db     503 SVKNGTYDYPKYSEESKLNRGKIDGVKLE 531


RESULT 35
AAE04952
ID   AAE04952 standard; Protein; 572 AA.
XX
AC   AAE04952;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Influenza virus A/Texas/36/91 recombinant haemagglutinin (rHA).
XX
KW   Multivalent influenza vaccine; recombinant haemagglutinin; rHA;
KW   baculovirus expression system; virucide; fusion protein;
KW   61K protein.
XX
OS   Chimeric - Autographa californica nuclear polyhedrosis virus.
OS   Chimeric - Influenza virus type A.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1..18
FT                    /note= "Autographa californica nuclear polyhedrosis virus
FT                    (AcNPV) 61K protein signal peptide"
FT   Protein          19..572
FT                    /note= "Influenza virus A/Texas/36/91 mature HA"
XX
PN   US6245532-B1.
XX
PD   12-JUN-2001.
XX
PF   09-OCT-1998;   98US-0169027.
XX
PR   30-MAY-1995;   95US-0453848.
PR   13-SEP-1993;   93US-0120607.
XX
PA   (PROT-) PROTEIN SCI CORP.
XX
PI   Smith GE,  Volvovitz F,  Wilkinson BE,  Voznesensky AI,  Hackett CS;
XX
DR   WPI; 2001-407272/43.
DR   N-PSDB; AAD09587.
XX
PT   Expressing a protein e.g. recombinant influenza virus hemagglutinin
PT   comprising using a vector encoding a polypeptide comprising a
PT   baculovirus signal peptide and a baculovirus expression system is
PT   useful as a multivalent influenza vaccine -
XX
PS   Claim 1; Column 45-48; 51pp; English.
XX
CC   The present invention relates to a method for expressing an exogenous
CC   protein in a baculovirus expression system which comprises using a vector
CC   encoding a polypeptide comprising a baculovirus signal peptide operably
CC   linked to a heterologous amino acid sequence. The method is especially
CC   useful for preparing a protein which may be used to make a multivalent
CC   influenza vaccine based on a mixture of recombinant haemagglutinin
CC   (HA) antigens cloned from influenza viruses having epidemic potential.
CC   The recombinant haemagglutinin proteins are full length,
CC   uncleaved (HAO) glycoproteins including both the HA1 and HA2 subunits
CC   (HAO) purified under non-denaturing conditions. The use of recombinant
CC   DNA (rDNA) technology to produce influenza vaccine offers several
CC   advantages, e.g., a recombinant DNA influenza vaccine can be produced
CC   under safer and more stringently controlled conditions; propagation with
CC   infectious influenza in eggs is not required; recombinant haemagglutinin
CC   (rHA) protein can be more highly purified, purification procedures for
CC   rHA do not have to include virus inactivation or organic extraction of
CC   viral membrane components, production of HA via rDNA technology provides
CC   an opportunity to avoid the genetic heterogenicity which occurs during
CC   the adaptation and passage through eggs, which should make it possible to
CC   better match vaccine stains with influenza epidemic stains, resulting in
CC   improved efficacy. The present sequence is recombinant haemagglutinin
CC   (rHA) protein comprising Autographa californica Nuclear Polyhedrosis
CC   Virus (AcNPV) 61K protein signal sequence linked
CC   to Influenza virus A/Texas/36/91 mature HA protein.
XX
SQ   Sequence   572 AA;

  Query Match             4.5%;  Score 105.5;  DB 22;  Length 572;
  Best Local Similarity   22.3%;  Pred. No. 4;
  Matches   60;  Conservative   37;  Mismatches   87;  Indels   85;  Gaps   14;

Qy     195 FGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSS--YKDYMDTQIPALPKFAKRFSLM 252
           || |       || ||  :|       :    | ||  :::     |   ||: :   |
Db     286 FGSG-------IITSNASMDECDAKCQTPQGAINSSLPFQNVHPVTIGECPKYVRSTKLR 338

Qy     253 VV----------QRLLATVAGYVDTPW---YKKWY------------------MKLKN 279
           :|          : |   :||:::  |      ||                      |
Db     339 MVTGLRNIPSIQSRGLFGAIAGFIEGGWTGMIDGWYGYHHQNEQGSGYAADQKSTQNAIN 398

Qy     280 FMVNRV----------FIPTKKFFNKEIREPSKALKEKVSTDTKD----------LFENK 319
           : |:|          |    | ||| :    : | :||     |          | ||
Db     399 GITNKVNSVIEKMNTQFTAVGKEFNK-LERRMENLNKKVDDGFLDIWTYNAELLVLLEN- 456

Qy     320 IGQGTVDFFNKEIRDPSKALKEKVSNDAKDLFENKIGQGTVDF---INNEIRDPSKALIR 376
           |: |:|| :  :::  : :| :: |:||:     || |  :|    |||       :
Db     457 -GR-TLDFHDSNVKNLYEKVKSQLKNNAKE-----IGNGCFEFYHKCNNE-------CME 502

Qy     377 KVSTGAEDL----FENKIGQGTVDFINNE 401
           |  |  |      |:|: :| :| :  |
Db     503 SVKNGTYDYPKYSEESKLNRGKIDGVKLE 531


RESULT 36
```

```
AAU33839
ID   AAU33839 standard; Protein; 465 AA.
XX
AC   AAU33839;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #115.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2..
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS51698.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 5335; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   465 AA;

  Query Match             4.5%;  Score 105;  DB 22;  Length 465;
  Best Local Similarity   19.9%;  Pred. No. 3.3;
  Matches   74;  Conservative   59;  Mismatches  124;  Indels  114;  Gaps   19;

Qy      99 QLEAAFILFKESDANPANSTEKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIEN 158
           |:||   | :| :  |       |  |:|               |  |   |   |:
Db      19 QIEAVLTLLEEKNTVP-------FIARYR----------------KEQTGGLDEVQIKQ 54

Qy     159 FASRYLYMATLYYK---TYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDI 215
                 | ||  |  :      |::: |                        ::   |  ||

Db      55 IDDEYQYMVNLQKRKEEVIKNIEQQG------------------------LLTEELKKDI 90

Qy     216 GTEHSVSRLQHITSSYKDYMDTQIPALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYM 275
              :: : |:: :   :|    |:       |||  |      ||             :|
Db      91 LKQNKLQRVEDLYRPFKQKKKTRATE----AKRKGL----EPLAI-------------WM 129

Qy     276 KLKNFMVNRVFIPTKKFFNKEIREPSKALK-------EKVSTDTKDLFENKIGQGTVDFF 328
           | :   |: :    ::| |:|::    |:|       |::| : |  :  ||     | :
Db     130 KARKHEVS-IEEKAQQFINEEVQSVEDAIKGAQDMIAEQISDNPK--YRTKI---LKDMY 183

Qy     329 NKEIRDPSKALKEKVSNDAKDLFE---------NKIGQGTVDFINNEIRDPSKALIRKV- 378
           :: :   |   |:| : | | :||          :|    |  :|    :  | |  |
Db     184 HQGVLTTS---KKKNAEDEKGIFEMYYAYSEPIKRIANHRVLAVNR--GEKEKVLSVKFE 238

Qy     379 --STGAED-LFENKIGQGTVD--FINNEIRDPSKALI-----RKVYTEADDLFENKIGQG 428
             :|  || :  :|    |:  :|   |:|  | ||     |::: :  :  ||
Db     239 FDTTSVEDFIARQEINHNNVNRSYILEAIKDSLKRLIVPSIEREIHADLTEKAENH---- 294

Qy     429 TVDFINKEIRD 439
           :|  :: :|:
Db     295 AIDVFSENLRN 305


RESULT 37
AAY20015
ID   AAY20015 standard; Protein; 697 AA.
XX
AC   AAY20015;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, t373.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   N-PSDB; AAX61712.
XX
PT   New isolated Borrelia burgdorferi nucleic acids  - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 12; Page 161; 275pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   697 AA;

  Query Match             4.5%;  Score 105;  DB 20;  Length 697;
```

```
  Best Local Similarity   22.0%;  Pred. No. 5.7;
  Matches   96;  Conservative   55;  Mismatches  143;  Indels  142;  Gaps   24;

Qy      98 YQLEAAFILFKESDANPANSTEKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATD-- 155
           |    ||   |:   |      |:| |:|    | :   |  :  |: :   : |  |
Db      58 YLEHMAFNGTKDYPGNSIVDVLKKFGMQFGADINAATSFDFTYYRLDLSDGNNKDEIDES 117

Qy     156 ---IENFASRYLYM-----------------ATLYYKTYTNVDEFGASFFNKLSFTTGL 194
              : |:||:  :|                 |   : |  :|:|          |:|
Db     118 INILRNWASQISFMKEEIDLERNIIIEEKKLGETYPGRIYEKMDKF---------LTSG- 167

Qy     195 FGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALPKFAKRFSLMVV 254
                   :     |  || |  :   |                 |:  |:|
Db     168 ----------SLYEFRSP--IGLEEQILSFQ----------------PEDFKKF----- 193

Qy     255 QRLLATVAGYVDTPWYKKWYM-KLKNFMVNRVFIP------TKKFFNKEIREPSKALKE- 306
                       |:|||  :| : :|     |        || |    : |:  :||
Db     194 --------------YRKWYRPELASVIVVGDIDPIEIEEKIKKQF-VSWKNPTDKIKEV 237

Qy     307 KVSTDT--KDLF----ENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKD-----LFENKI 355
           |||  |   || |    : ::|: :: || ||| :  |  |: : |  |      ||||:
Db     238 KVSLDVELKDKFLLLEDLEVGEPSLMFFKKEIINFVKT-KDDLLNAIKKSLLAALFENRF 296

Qy     356 GQ----GTVDFINNEIRD------PSKALIRK----------VSTGAEDLFE-----NKI 390
             :    |  | |   :|       :  :: |          :: | :| |       |
Db     297 SELKTAGVKQFKNVSNKDFFSFKSDNNTIVAKSISLNFNPDHLNEGIQDFFYELERIRKF 356

Qy     391 G--QGTVDFINN------EIRDPSKALIRKVYTEA--DDLFENKIGQGTVDFINKEIRDP 440
           |  || :: : :       |:|   |  | |  : |   || |  |  |:  |   |  |
Db     357 GFTQGELEKVRSQFYKSLELR---KKNINKTNSWAIFQDLIEIAI-NGSNKFDMNEYCDL 412

Qy     441 SKALIRKVSTEADNLL 456
           |   : |:  :  | |
Db     413 SFQYLEKIDLKTINNL 428


RESULT 38
AAU36810
ID   AAU36810 standard; Protein; 716 AA.
XX
AC   AAU36810;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #980.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54669.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12403; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   716 AA;

  Query Match             4.5%;  Score 105;  DB 22;  Length 716;
  Best Local Similarity   19.9%;  Pred. No. 5.9;
  Matches   74;  Conservative   59;  Mismatches  124;  Indels  114;  Gaps   19;

Qy      99 QLEAAFILFKESDANPANSTEKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATDIEN 158
           |:|)   | :| :  |       |  |:|                 |  |   |   |:
Db      19 QIEAVLTLLEEKNTVP-------FIARYR----------------KEQTGGLDEVQIKQ 54

Qy     159 FASRYLYMATLYYK---TYTNVDEFGASFFNKLSFTTGLFGWGIKRALKQIIRSNLPLDI 215
                | ||  |  :      |::: |                         ::   |  ||
Db      55 IDDEYQYMVNLQKRKEEVIKNIEQQG-----------------------LLTEELKKDI 90

Qy     216 GTEHSVSRLQHITSSYKDYMDTQIPALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYM 275
             :: : |:: :   :|    |:       |||  |      ||              :|
Db      91 LKQNKLQRVEDLYRPFKQKKKTRATE----AKRKGL----EPLAI------------WM 129

Qy     276 KLKNFMVNRVFIPTKKFFNKEIREPSKALK-------EKVSTDTKDLFENKIGQGTVDFF 328
           | :   |: :    ::| |:|::    |:|       |::| : |  :  ||     | :
Db     130 KARKHEVS-IEEKAQQFINEEVQSVEDAIKGAQDMIAEQISDNPK--YRTKI---LKDMY 183

Qy     329 NKEIRDPSKALKEKVSNDAKDLFE---------NKIGQGTVDFINNEIRDPSKALIRKV- 378
           :: :   |   |:| : | | :||           :|    |  :|    :  | |  |
Db     184 HQGVLTTS---KKKNAEDEKGIFEMYYAYSEPIKRIANHRVLAVNR--GEKEKVLSVKFE 238

Qy     379 --STGAED-LFENKIGQGTVD--FINNEIRDPSKALI-----RKVYTEADDLFENKIGQG 428
             :|  || :  :|    |:  :|   |:|  | ||     |::: :  :  ||
Db     239 FDTTSVEDFIARQEINHNNVNRSYILEAIKDSLKRLIVPSIEREIHADLTEKAENH---- 294

Qy     429 TVDFINKEIRD 439
            :|  :: :|:
Db     295 AIDVFSENLRN 305


RESULT 39
AAY20014
ID   AAY20014 standard; Protein; 719 AA.
XX
AC   AAY20014;
XX
DT   19-JUL-1999  (first entry)
XX
DE   B. burgdorferi antigenic protein, f373.aa.
XX
KW   Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS   Borrelia burgdorferi.
```

```
XX
PN   WO9859071-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12718.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Choi GH,  Erwin AL,  Hanson MS,  Lathigra R;
XX
DR   WPI; 1999-189980/16.
DR   N-PSDB; AAX61711.
XX
PT   New isolated Borrelia burgdorferi nucleic acids  - used to develop
PT   products for the diagnosis, prevention and treatment of diseases
PT   caused by Borrelia, particularly Lyme disease
XX
PS   Claim 12; Page 161; 275pp; English.
XX
CC   This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC   invention, which is suitable for use in a vaccine. The Bb polypeptides
CC   can be used in vaccines for eliciting protective antibodies to members of
CC   the Borrelia genus, particularly for the use against Lyme disease in
CC   humans and animals. They can be used for preventing or attenuating an
CC   infection caused by a member of the Borrelia genus. The products can also
CC   be used for detection of members of the Borrelia genus.
XX
SQ   Sequence   719 AA;

  Query Match             4.5%;  Score 105;  DB 20;  Length 719;
  Best Local Similarity   22.0%;  Pred. No. 6;
  Matches   96;  Conservative   55;  Mismatches  143;  Indels  142;  Gaps   24;

Qy      98 YQLEAAFILFKESDANPANSTEKRFWMRFRRGKNHSYFHDLVFNLLEKNVTRDADATD-- 155
           |    ||   |:   |       |:| |:|    | :   |  :  |: :  : |  |
Db      80 YLEHMAFNGTKDYPGNSIVDVLKKFGMQFGADINAATSFDFTYYRLDLSDGNNKDEIDES 139

Qy     156 ---IENFASRYLYM-----------------ATLYYKTYTNVDEFGASFFNKLSFTTGL 194
             : |:||:  :|                  |   : |  :|:|          |:|
Db     140 INILRNWASQISFMKEEIDLERNIIIEEKKLGETYPGRIYEKMDKF---------LTSG- 189

Qy     195 FGWGIKRALKQIIRSNLPLDIGTEHSVSRLQHITSSYKDYMDTQIPALPKFAKRFSLMVV 254
                      :     |  || |  :   |                 |:  |:|
Db     190 ----------SLYEFRSP--IGLEEQILSFQ----------------PEDFKKF----- 215

Qy     255 QRLLATVAGYVDTPWYKKWYM-KLKNFMVNRVFIP------TKKFFNKEIREPSKALKE- 306
                     |:|||  :| : :|     |       || |    : |:  :||
Db     216 --------------YRKWYRPELASVIVVGDIDPIEIEEKIKKQF-VSWKNPTDKIKEV 259

Qy     307 KVSTDT--KDLF----ENKIGQGTVDFFNKEIRDPSKALKEKVSNDAKD-----LFENKI 355
           |||  |   || |    : ::|: :: || ||| :  |  |: : |  |      ||||:
Db     260 KVSLDVELKDKFLLLEDLEVGEPSLMFFKKEIINFVKT-KDDLLNAIKKSLLAALFENRF 318

Qy     356 GQ----GTVDFINNEIRD------PSKALIRK----------VSTGAEDLFE-----NKI 390
            :    |   | |   :|       :  :: |          :: | :| |        |
Db     319 SELKTAGVKQFKNVSNKDFFSFKSDNNTIVAKSISLNFNPDHLNEGIQDFFYELERIRKF 378

Qy     391 G--QGTVDFINN------EIRDPSKALIRKVYTEA--DDLFENKIGQGTVDFINKEIRDP 440
           |  || :: : :      |:|   |  | |  : |   || |  |  |:  |   |  |
Db     379 GFTQGELEKVRSQFYKSLELR---KKNINKTNSWAIFQDLIEIAI-NGSNKFDMNEYCDL 434

Qy     441 SKALIRKVSTEADNLL 456
           |   : |:  :  | |
Db     435 SFQYLEKIDLKTINNL 450
```

```
RESULT 40
AAG79241
ID   AAG79241 standard; Protein; 719 AA.
XX
AC   AAG79241;
XX
DT   03-JAN-2002  (first entry)
XX
DE   Amino acid sequence of S-layer protein of C. difficile strain 630.
XX
KW   Surface layer protein; S-layer protein; pseudomembranous colitis; PMC;
KW   cell wall protein; gastrointestinal illness; abscess; wound infection;
KW   osteomyelitis; urogenital tract infection; septicemia; peritonitis;
KW   pleuritis.
XX
OS   Clostridium difficile.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 105
FT                   /note= "Asp encoded by GTT"
XX
PN   WO200173040-A1.
XX
PD   04-OCT-2001.
XX
PF   23-MAR-2001; 2001WO-GB01305.
XX
PR   24-MAR-2000; 2000GB-0007263.
XX
PA   (UNLO ) IMPERIAL COLLEGE SCI TECHNOLOGY & MED.
XX
PI   Fairweather NF,  Calabi E;
XX
DR   WPI; 2001-616508/71.
DR   N-PSDB; AAI65840.
XX
PT   Novel polypeptides and polynucleotides of cell wall proteins of
PT   Clostridium difficile especially S-layer cell wall protein useful for
PT   preventing and treating the infection caused by the bacteria  -
XX
PS   Claim 1; Page 56; 62pp; English.
XX
CC   The present sequence represents a surface layer (S-layer) protein of
CC   Clostridium difficile. The S-layer proteins are the predominant cell
CC   wall protein. There are two distinct S-layer proteins in C. difficile,
CC   a 45 kDa and 36 kDa protein. S-layer polypeptides and polynucleotides
CC   are useful for treating and/or preventing a disease associated with
CC   C.difficile infection in a subject. Such diseases include
CC   pseudomembranous colitis (PMC) in humans characterized by diarrhoea, a
CC   severe inflammation of the colonic mucosa, and formation of
CC   pseudomembranes that are composed of fibrin, mucus, necrotic epithelial
CC   cells and leukocytes; gastrointestinal illness, abscesses, wound
CC   infections, osteomyelitis, urogenital tract infections, septicemia,
CC   peritonitis, and pleuritis.
XX
SQ   Sequence   719 AA;

  Query Match             4.5%;  Score 105;  DB 22;  Length 719;
  Best Local Similarity   21.8%;  Pred. No. 6;
  Matches  106;  Conservative   75;  Mismatches  185;  Indels  120;  Gaps   26;

Qy       8 GDVTK--TLLAASESVDSAANAYMINS-------DMSDYLSAVSDNFAERICSQVPKGSN 58
           | |||  | || | ::    |  :||:       | | | ||  :| |:|      : |
Db     252 GFVTKDDTDLAKSGTI----NVRVINAKEESIDIDASSYTSA--ENLAKRYVFDPDEIS- 304

Qy      59 CSASVSAYMSRCAKQDCLTLQSLKYPLEAKYQPLTLPDPYQLEAAFILFKESDANPANST 118
             || :  | |:     :|   :  ||| :  |:  :||        :|   :  |
Db     305 -----EAYKAIVALQNDGIESNLVQLVNGKYQVIFYPEGKRLETK----SANDTIASQDT 355

Qy     119 EKRFWMRFRRGKN-HSYFHDL-VFNLLEKNVTRDADATDIENFASRYLYMATLYYKTYTN 176
             :  ::  : |:   |  ||  :|    ||   |    ||      : ::: ||    |
Db     356 PAKVVIKANKLKDLKDYVDDLKTYNNTYSNVVTVAGEDRIET----AIELSSKYY----N 407
```

```
Qy     177 VDEFGA---------SFFNKLSFTTGLFGWGI--KRALKQIIRSNLPLDIGTEHSVSRLQ 225
           |:  |              |   ||    :  ::    :: |   ||   :  : |:
Db     408 SDDKNAITDKAVNDIVLVGSTSIVDGLVASPLASEKTAPLLLTSKDKLDSSVKSEIKRVM 467

Qy     226 HITSSYKDYMDTQIPALPKFAKRFSLMVVQRLLATVAGYVDTPWYKKWYMKLKNFMVNRV 285
           :: |      || |    |               :|| |::   |    :|||     :
Db     468 NLKS------DTGINTSKK-------------VYLAGGVNS-ISKDVENELKN-----M 501

Qy     286 FIPTKKFFNKEIREPSKALKEKVSTDTKDLFENKIGQGTVDFFN-----KEIRD------ 334
           :   :   ::  | | |: :::  |    |    | |  |  :      :::|
Db     502 GLKVTRLSGEDRYETSLAIADEIGLDNDKAFVVG-GTGLADAMSIAPVASQLKDGDATPI 560

Qy     335 ---PSKALKEKVSNDAKDLFENKIGQGTVDFINNEIRDPSKALIRKVSTGAEDLFENKIG 391
               ||  :::|:|||      :|   || |  :           ||   |:  ::  |
Db     561 VVVDGKA--KEISDDAKSF----LGTSDVDIIGGK---------NSVSKEIEESIDSATG 605

Qy     392 QGTVDFINNEIRDPSKALIRKVYTEADDLFENKIGQGTVDFINKEIRDPSKALIRKVSTE 451
           : | | |: : |  : | : |     || | :  |:    |: |:            ||:
Db     606 K-TPDRISGDDRQATNAEVLK----EDDYFTD--GEVVNYFVAKD----------GSTK 647

Qy     452 ADNLLE 457
           | |::
Db     648 EDQLVD 653
```

Search completed: October 19, 2002, 07:49:05
Job time : 69 secs